Department of Philosophy, Logic and Scientific Method
London School of Economics and Political Science

The Epistemology of AI-driven Science: The Case of AlphaFold

Daria Zakharova

Abstract

The success of AlphaFold, an AI system that predicts protein structures, poses a challenge for traditional understanding of scientific knowledge. It operates opaquely, generating predictions without revealing the underlying principles behind its predictive success. Moreover, the predictions are largely not empirically tested but are taken at face value for further modelling purposes (e.g. in drug discovery) where experimentation takes place much further down the line. The paper presents a trilemma regarding the epistemology of AlphaFold, whereby we are forced to reject one of 3 claims: (1) AlphaFold produces scientific knowledge; (2) Predictions alone are not scientific knowledge unless derivable from established scientific principles; and (3) Scientific knowledge cannot be strongly opaque. The paper argues that AlphaFold's predictions function as scientific knowledge due to their trustworthiness and functional integration into scientific practice. The paper addresses the key challenge of strong opacity by drawing on Alexander Bird's functionalist account of scientific knowledge as irreducibly social, and advances the position against *individual* knowledge being necessary for the production of *scientific* knowledge. It argues that the implicit principles used by AlphaFold satisfy the conditions for scientific knowledge, despite their opacity. Scientific knowledge can be strongly opaque to humans, as long as it is properly functionally integrated into the collective scientific enterprise.

Introduction

Artificial intelligence (AI) has drastically transformed many domains of science, with profound implications for the production and validation of scientific knowledge. One striking case is that of AlphaFold, an AI system designed to predict protein structures. The problem of predicting how a protein's sequence of amino acids determines its three-dimensional structure—the protein folding problem—has historically been one of the most difficult and significant challenges in biological sciences. AlphaFold's success in producing a data bank of close to all protein structures known in nature has revolutionized the biological sciences. Demis Hassabis and John Jumper recently received the Nobel Prize[1] for their work on AlphaFold, amplifying the debate about whether they have succeeded in solving the protein-folding problem (Al-Janabi 2022)[2][3].

AlphaFold's unprecedented success in producing reliable predictions of protein structures presents a fundamental challenge for epistemology of science: it operates as a black box, generating predictions without revealing the underlying mechanism or principles it uses. This opacity raises critical questions about the nature and the production of scientific knowledge in the age of AI-driven science. Can we claim that AlphaFold generates *new scientific knowledge* when its processes are inaccessible to human understanding? Are predictions alone, even highly reliable ones, sufficient to constitute scientific discovery? Or is it simply a tool that does not, by itself, generate scientific knowledge?

This paper investigates whether AlphaFold's predictions are scientific knowledge. I propose a trilemma that confronts us with the challenge of accommodating opaque AI systems like AlphaFold within our broader understanding of how science generates knowledge. Can we continue to rely on traditional empiricist frameworks, or do we need to develop new epistemological models that better reflect the realities of AI-driven research? In the first section, I present the protein-folding problem and the general workings of AlphaFold. The second section presents a trilemma regarding the epistemology of AlphaFold. I argue against an empiricist account in order to defend the claim that AlphaFold generates scientific knowledge, and that knowledge can be strongly opaque. In the third section I respond to some objections by drawing on apparent similarities with older challenges to reliabilism. I conclude by

[1] https://www.nature.com/collections/edjcfdihdi
[2] https://magazine.hms.harvard.edu/articles/did-ai-solve-protein-folding-problem
[3] https://www.scientificamerican.com/article/one-of-the-biggest-problems-in-biology-has-finally-been-solved/

discussing an implication of the thesis that AlphaFold generates scientific knowledge, namely that scientific knowledge can be strongly opaque to human scientists.

1. Background

1.1. Protein folding and AlphaFold

A protein is a sequence of amino acids. When interacting with any environment, for instance when put in water, the proteins interact with other molecules by *folding* the amino acid string into a three-dimensional structure. This 3D native structure of the protein is thought to be encoded in its 1D amino acid string, but the mechanism and the physical law behind this native encoding is not well understood, with several hypotheses having been proposed (Dill and McCallum 2012; Dill et al. 2008). The *protein folding problem* is the question of how a protein's amino acid sequence dictates its three-dimensional atomic structure (Dill et al. 2008). There is so far no known general mechanism or law-like rule that explains how proteins fold in nature. As proteins perform many important functions in biology and biochemistry, predicting their folds has been one of the most important and most difficult problems in biological sciences. Since (cheaper) computational methods for modelling proteins became available and increasingly successful, predicting protein structures to accelerate drug discovery became a major objective in computational biology (ibid.). Despite the advances in machine learning in the early 2010s, the Protein Data Bank (PDB), which is the primary repository for protein structures, had approximately 170,000 structures by 2020. These structures were determined still using primarily experimental methods such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM). The process of determining protein structures experimentally has historically proven to be extremely time-consuming, costly, and not feasible for every protein, leading to a substantial gap between the number of known 1D protein sequences and their corresponding 3D structures. In 2019 Deep Mind announced *AlphaFold*, an artificial intelligence system which can successfully predict protein structures, and in 2021 AlphaFold 2.0 came out (Jumper et al. 2021), Then in 2024 AphaFold 3.0 has been released, with the newest version able to predict DNA, RNA and ligand structures, all essential to further accelerating drug discovery[4] (Abramson et al. 2024). AlphaFold 2.0, the powerful iteration based on a transformer

[4] https://www.technologyreview.com/2024/10/09/1105335/google-deepmind-wins-joint-nobel-prize-in-chemistry-for-protein-prediction-ai/

architecture is the one I will be primarily focusing on in this paper. As is typical of deep learning algorithms, the system is a black box: it generates output without making the internal processes for reaching specific decisions known to the user or engineer. It does not 'explain' which mechanism it uses for protein folding. AlphaFold can predict protein folds with an atomic level of precision and has been integrated into our newest scientific developments in biology, chemistry, and medicine (Yang, Zeng and Chen 2023). As of 2024, AlphaFold has predicted the structures of over 200 million proteins. This comprehensive database includes nearly all known protein structures from a wide range of organisms, including plants, bacteria, animals, and viruses. The massive expansion from its initial release has significantly impacted the scientific community, accelerating research and innovation in fields such as drug discovery, molecular biology, and biotechnology[5]. Protein structure prediction on this scale was a major breakthrough in science. In its citation for the 2024 Nobel Prize in Chemistry, the Nobel Committee wrote that "Demis Hassabis and John Jumper have developed an AI model to *solve* a 50-year-old problem: predicting proteins' complex structures. These *discoveries* hold enormous potential"[6] (italics added).

1.2. The problem

AlphaFold is said to have discovered hundreds of millions of protein structures (Jumper et al. 2021), effectively granting the novel *predicted* structures the status of *discovered* ones in the public discourse. A subset of its novel predictions has been empirically confirmed, proving the system to be highly reliable. All these new protein structures have *not* led human scientists to discover a principle / mechanism of protein folding, such that would allow them to either propose an explanation of how AlphaFold could have done it or to independently advance a theory of protein folding themselves. We did not learn why and how proteins fold the way they do, despite having a vast number of new predictions, that we are actively using in advancing our science, that mostly do not even undergo experimental confirmation but instead can be readily used from AlphaFold's comprehensive database. Despite its architecture not being opaque in the most direct sense to its creators, AlphaFold appears to have solved a problem humans could not and *still cannot* solve. Therefore, at least in epistemological terms, I propose that it produced these highly reliable predictions by an opaque method, having been given all the knowledge about proteins currently available to us to learn from as its training dataset.

---

[5] (SciTechDaily) (MIT Technology Review) (Enterprise Technology News and Analysis)
[6] Press release. NobelPrize.org. Nobel Prize Outreach AB 2024. Tue. 5 Nov 2024. https://www.nobelprize.org/prizes/chemistry/2024/press-release/

Furthermore, most scientists using AlphaFold do not take its prediction and first do an experiment to check if a particular protein actually folds accordingly. Instead, they take the given structure and use it to further model whatever their objective is. There is normally no experimentation involved until a much later stage of the process in which AlphaFold is involved. For instance, in drug discovery research, AlphaFold can be employed for determining structures that could be useful for a certain purpose (looking for a specific target), together with computational chemistry platforms such as Chemistry42, and biocomputational generative platforms such as PandaOmics. Chemistry42 has been used to generate the molecules based on the structures predicted by AlphaFold. When the novel target of interest is identified – these molecules are tested in biological assays (Ren et al. 2023). While there is laboratory testing of the molecules selected for the research purpose, it happens much later, after AlphaFold is employed to generate a pool of possible targets of interest, on the basis of which further generated structures are produced, and only a subset of those is then biologically tested. AlphaFold is used to find prospective targets, where its outputs seem to be held as trustworthy enough for scientists to use them for further modelling and novel targets search, without having to confirm the structures empirically (which would not be practically feasible anyway and defy the purpose of AlphaFold use).

This leads us to an arguably strange situation from an epistemological point of view. Determining protein structures has been both one of the most important and one of the most difficult problems in biological sciences, which now appears to have been 'solved' by an opaque method. Since we use these novel protein structures ubiquitously in our cutting-edge science – we should normally say that these structures are part of scientific knowledge, of what is 'known to science'. On the other hand, the worry is that computational predictions, mostly not empirically tested, and for which we have no underlying theory or law, cannot qualify for scientific knowledge or genuine discovery, prompting questions for epistemology of AI-driven science, e.g. whether equating prediction with discovery in this context is a mistake. Highlighting this as a precedent is the long tradition of Nobel Prizes being awarded for empirically tested discoveries, not for novel techniques or technologies. Consider that Jumper and Hassabis may be said to have cracked the principle behind protein folding but it would seem that neither them nor other scientists can hardly claim to *themselves have the knowledge* of this principle.

The central question is whether AlphaFold generates scientific knowledge. And if it does, another question is who *has* this knowledge – science, scientists, AlphaFold? In the next section I set out a puzzle arising from the idea that AlphaFold produces scientific knowledge.

1. Does AlphaFold generate scientific knowledge? A trilemma.

Given these developments, we need to accommodate AlphaFold, and similar transformer AI systems, in our conception of how scientific knowledge is produced. I propose to approach the epistemology of AlphaFold in form of a trilemma. I first posit three claims which arise from the use of AlphaFold and similar AI systems in scientific practice:

1. AlphaFold produces scientific knowledge.
2. Empirically unconfirmed predictions alone are not scientific knowledge unless derivable in an appropriate way from supporting theories, laws or mechanisms that are scientific knowledge.
3. Scientific knowledge cannot be strongly opaque.

First, I explain how each of the claims is plausible, and we normally have good reasons to accept them:

1. **AlphaFold produces scientific knowledge.** We see that the predictions AlphaFold has generated so far are highly reliable and ubiquitously used by scientists. Normally, scientists do not even interact with AlphaFold itself, instead they use the database of the novel protein structures it has produced, and just take the ones they need from there. Moreover, there is usually no experiment they conduct to confirm the structure before they use it in their work, as most work is done with further modelling, e.g. in drug discovery and development. AlphaFold produced a repository of claims (about protein structures). If a repository of scientific claims is regarded with good reason as authoritative and trustworthy in scientific practice, it should (at least at face value, absent any reason to believe the trust is misplaced) be regarded as scientific knowledge.
2. **Empirically unconfirmed predictions alone are not scientific knowledge – unless derivable in an appropriate way from supporting theories, laws or mechanisms that are scientific knowledge.** Sometimes predictions are plausible cases of scientific knowledge (e.g. plausibly, it is known that Halley's comet will return in 2061). However, these are cases where there is a known theory, a law or a mechanism that produces the prediction.

One might find this proposition too demanding, thinking of examples such as "it is scientific knowledge that paracetamol reduces fever, even though there is no knowledge of *how* exactly it does that”. However, it is *not* that a *generalization* cannot be scientific knowledge unless an underlying mechanism is known. It's rather that a prediction about a specific instance cannot be scientific knowledge unless (at minimum) the supporting generalizations are known.

One could retort that we *do* in fact know the supporting generalization – the generalization is that AlphaFold is x% reliable. But this is akin to someone claiming that they know that they will not win the lottery because only one in a million tickets wins. I might know the odds of winning, but I cannot, on the basis of this, claim knowledge of *whether I will win this particular time I play*. Similarly, I may know paracetamol works in x% of cases, but it is not scientific knowledge that it will work *in my particular case* until its effects on me have been observed. And, in the case of AlphaFold, a prediction of a specific protein structure, before confirmation and without a known mechanism or theory behind it, does not seem to give us scientific knowledge.

To see the absurdity of taking a predictor with a good track record to be generating scientific knowledge of the truth of its predictions, imagine a “science guru” who has a good track record of predicting future Nobel Prize-worthy discoveries. It does not seem plausible to equate this oracle’s predictions with discoveries, equating the predictions/bets with knowledge. If this were correct, we should be giving *them* the Nobel Prize, since the relevant knowledge has already been achieved.

The restriction to “scientific” knowledge is important. One can imagine scenarios where reliable predictions may meet the standards for *everyday* knowledge despite the lack of any derivation from scientific knowledge (e.g. perhaps I can know that sun will rise and set tomorrow even if I know none of the relevant astronomical generalizations). But the prediction will be scientific knowledge only if the supporting generalization is too.

3. **Scientific knowledge cannot be strongly opaque.** To the extent that AlphaFold’s predictions are derived from a supporting theory, law or mechanism of protein folding, the supporting generalizations are “strongly opaque”. *Strong* opacity implies that no individual human knows the relevant theory, law or mechanism, and, moreover, no human has any way of accessing it. This can be distinguished from cases of *weak* opacity in which only a few experts possess the knowledge or in which the knowledge is difficult but possible to access.

> Why think scientific knowledge cannot be strongly opaque? If a black box falls out of the sky, it seems absurd to entertain the possibility that we can immediately treat the information it contains as scientific knowledge even if *no one* can access that information! If current AI systems make strongly opaque scientific knowledge possible, this is something without any obvious precedent.

While each claim is on its own plausible, the strong opacity of AlphaFold's implicit theory, law or mechanism of protein folding forces us to reject one of them. Which one should we reject?

An empiricist (my imagined critic) might want to reject (1) and defend (2) and (3). An empiricist I have in mind here is one that defends a more 'traditional', internalist empiricism (e.g. van Fraassen (1980, 1985) might represent), giving the central place of scientific inquiry to observation through some type of sense-data. For her, evidence is constituted through observation, which is itself tied to some form of belief or a mental state of the scientist. Empirical confirmation thus depends on the scientist knowing and on conducting experiments that ought to provide the empirical observation data. To what degree is aided observation allowed on this picture does not make a big difference for this case (e.g. one could adopt a permissive view following Shapere (1982) (Boyd and Bogen 2021)).

In the rest of this paper, I will defend (1) and (2) and reject (3). I will proceed by discussing the reasons to support and to reject each of the claims, ultimately arguing for a picture on which AlphaFold produces scientific knowledge even though part of this knowledge is strongly opaque.

2.1 Defending 1: AlphaFold produces scientific knowledge

Let's consider the first claim, beyond the prima facie motivating reasons for it I initially described above: AlphaFold produces scientific knowledge.

In first instance this claim is about the status of the direct output of AlphaFold – the predicted novel protein structures, most of which are not empirically confirmed but highly reliable. These can be understood as propositional in form and are themselves not opaque to human scientists, as they are available through a comprehensive AlphaFold database. The argument for or against the claim then depends on how we answer the question: Can a prediction (not a generalization but a prediction of e.g. a specific instance of a protein structure), not yet empirically confirmed, count as discovery / knowledge if the predicting system is known to be highly reliable? I intend to answer this question positively, and provide two main reasons below, thereby defending the claim of the Trilemma. An empiricist opponent, however, might want to argue as follows:

P1: Even when reliable, predictions are not discoveries (and therefore not scientific knowledge) before empirical confirmation.

P2: AlphaFold's individual predicted protein structures are reliable but not empirically confirmed.

C: Therefore, AlphaFold itself does not make discoveries or generate knowledge.

She therefore answers the question implied by the first premise negatively, thereby attacking the claim that AlphaFold generates knowledge.

It is important to clarify the grounds for skepticism here that an empiricist might raise. First, the claim concerns specifically scientific knowledge, and not 'knowledge in general' or 'everyday knowledge'. An empiricist might see the instance of confirmation of a prediction to be one that turns it into knowledge, which corresponds to an intuitive "prediction is confirmed – now we know!". An empirical confirmation, on this view, would differentiate between a prediction about novel protein folds and a discovery, where only the latter can be equated with scientific knowledge.

Contrary to this, I want to argue that a prediction can be knowledge before or in some cases without an empirical confirmation, and it is in fact part of normal scientific practice (with a caveat to which I will turn in the next section). To defend this claim, I illustrate a) how empirical confirmation is often missing for scientific predictions which effectively function as knowledge and b) the rather grey area of what exactly constitutes confirmation, especially as conceptualized by philosophers vs. by scientists. The latter point is related to the discussion of types of evidence accepted in modern science and whether they stand in a conflict with the stricter demands of empiricism among philosophers of science (Bird 2022).

Consider the first part of my argument, that empirical confirmation is not always necessary for a prediction to become knowledge. One way to think about this is rather trivial, that is when we consider that having a strong reliable prediction p gives us knowledge that p is highly likely to be true, so we can operate as if it is true, unless we gain further knowledge that lowers our credence in p. In AlphaFold's case – there is no problem in saying, at least on a reliabilist view (Goldman 1979, 1986), that the scientists can claim knowledge of individual predictions (therefore scientists can have individual knowledge of those novel protein structures), as reliabilism grants justification based on the reliability of the process that causes the belief in question (the structure of the protein is XYZ). In virtue of the reliability – they can be treated as knowledge unless e.g. our credence in their reliability gets lowered in light of some other evidence. However, one might say that scientists knowing the (propositional content of) prediction is not the same as scientists having a new piece of *scientific* knowledge – *that* would surely require the confirmation of the prediction first? However, science routinely has empirically unconfirmed predictions function as part of scientific knowledge. This most obviously concerns what an empiricist might regard as *unobservable objects* (Van Fraassen 1980), those for which no direct or even aided observation is either possible in principle or simply not yet gained by science.

An example of prediction that could be classified as a discovery before empirical confirmation was the discovery of the planet Neptune, which was predicted, and some would argue discovered through mathematical methods rather than direct observation. In 1846, irregularities in the orbit of Uranus led astronomers to predict the existence of another planet. Neptune was subsequently observed in the predicted position. While there can be a split in opinion about whether Neptune can be said to have been discovered before it was first observed visually, it only stresses the point about the complicated relationship between prediction, confirmation and discovery.

Some predictions are so robust and influential that they can be considered discoveries even before empirical confirmation, especially considering such predictions often drive research and experimental efforts. For instance, the prediction of antiparticles by Paul Dirac was considered groundbreaking and led to significant advancements in particle physics, even before the positron was experimentally discovered (Bird 2022).

As we see in the examples, and arguably in the usual scientific practice, scientific knowledge often includes and encompasses predictions without empirical confirmation, at least of the kind that implicitly involves sense-data observations of the predicted object or phenomenon itself. But what of the confirmation without sense-data observation? It could be argued that it still is required for the prediction to qualify for scientific knowledge.

Second part of the argument concerns the status and the type of the confirmation itself. What does it mean for a prediction to be confirmed, and which kind of confirmation is required? I argued above that a lot of scientific evidence towards strengthening or confirming predictions does not rely on sense-data observations. Scientific evidence in general often includes and heavily relies on simulations, models and inferences based on indirect observations and measurements (Bird 2022, 2010). Still, different accounts of confirmation may posit different requirements. AlphaFold for instance, may be conducting a kind of Bayesian confirmation, its architecture allowing for multistep / multilevel testing and updating. It is simulating new data on the basis of prior evidence, then testing it, then simulating again (there may be many steps involved in this process, as can be inferred from the AlphaFold architecture seen in Fig. 1). This process may be even viewed as a powerful simulation of an experiment, or rather of a great number of experiments, all run inside of the neural network (the opacity will be addressed in the next section). In general, despite varying possible metaphysical commitments of actual individual scientists, the ways evidence is generated and integrated to strengthen or disconfirm our scientific hypotheses is not restricted to sense-data and even aided observation. Confirming predictions has long been done by computers, often without involving any perceptual observation, and even when it did involve observation by a human – the role of perception was rather incidental and it can be easily substituted by a machine, as has been argued by Bird (2022, 159).

I have argued that predictions can sometimes constitute scientific knowledge before they are empirically confirmed. Second, even when there is confirmation behind a prediction in science – it very often falls short of involving sense-perception in a relevant sense, my empiricist critic might be committed to. In general, depending on the account of confirmation one can accept, and the way theory and prediction are connected, there are cases in which predictions themselves can be knowledge before they are empirically and / or directly confirmed.

As I have outlined the reasons to positively motivate the proposition that AlphaFold generates scientific knowledge in the previous section, the question for the empiricist remains on what to call AlphaFold's outputs, if *not* scientific knowledge. Some type of "quasi-knowledge" could be invoked here of course. However, if this "not quite knowledge" or "quasi-knowledge" has exactly the same functional role in science as knowledge (scientists treating AlphaFold's predictions of protein structures as actual protein structures), this distinction amounts to little more than paying lip service to traditional empiricism. It is therefore not really a distinction worth drawing. On this basis I would defend the claim that AlphaFold's predictions function as scientific knowledge, and therefore can be knowledge. The defense of this claim, however, is incomplete, without the discussion of how these predictions come about. The next section discusses this problem.

2.2. Defending 2: Empirically unconfirmed predictions alone are not scientific knowledge – unless derivable in an appropriate way from supporting theories, laws or mechanisms that *are* scientific knowledge

A novel prediction in science seems to require grounding in some known principle that gives this prediction reliability high enough for it to function as scientific knowledge, approximately in the sense I discussed in the previous section.

Consider an example from physics: When exactly have we discovered black holes and since when does science know they exist? The Nobel Committee states that it can only award the Nobel Prize for empirically tested work. The Nobel Prize for the work on black holes was awarded to Roger Penrose as recently as 2020. The first photograph of a black hole was taken in 2019. At the same time, the existence of black holes was (arguably) scientific knowledge for decades, long time before we had anything like a direct empirical confirmation[7] of their existence. The existence of black holes has been a major prediction of general relativity. This prediction has gained credibility through evidence that both supports the theory and reinforces the inference of black holes' existence, even though these inferences were not based on direct observation of the black holes themselves. Since black holes were central to the GR theory, physicists have been doing physics that posits the existence of black holes since at least the 1980s, effectively treating their prediction as part of "what is known to physics" – scientific knowledge. At which point in that story did the prediction turn into discovery can still be treated as a question with no definitive answer / no fact of the matter involved. However, we can say that the existence of black holes was known to science before we had empirical observations of black holes. We can claim to have had knowledge before empirical observation because they were not mere predictions, but predictions made by a very strong theory – a predictive framework known to be highly reliable. There are examples in history of science where correct predictions were made by an erroneous theory (Bird 2022, 53; O'Connell 2018). We can adopt Dellsen's (2016, 72) notion of scientific progress, on which correct, non-accidental predictions amount to knowledge as *understanding*. This implies that scientists' comprehension allows to make accurate predictions in related but non-actual situations (Bird 2022, 64). One could interpret his view to entail that correct predictions must be produced by a reliable method, rather than by chance. Bird (2022, 64) notes that correct prediction, even a reliably correct

[7] That said, the issue of direct vs indirect confirmation of black holes is complicated, even in regard to the current work on and the photographic evidence.

prediction, is too weak to count as understanding, which on his view is a subspecies of knowledge – knowledge of explanation. Understanding may be a subspecies of knowledge, e.g. it can be taken as necessary for knowledge of an explanation. For the present purposes, I am not interested in the ‘explanation’ or ‘understanding’ part of scientific knowledge, only in the power of reliable predictions functioning as part of knowledge if they are produced non-accidentally.

Even in case an empiricist accepts that empirical confirmation of a prediction is not always necessary for it to be knowledge, clearly a prediction alone is still insufficient for knowledge, if nothing can be said of the principles it is either derived from or which give it a high probability of being true. While it is contentious to assume fundamental principles or laws in biology, we can speak of mechanisms that can be known and it can be articulated how these mechanisms made the prediction possible and reliable.

Relevant to the case of protein folding predictions may be an example from John Norton’s “The Material Theory of Induction” (2021): Knowing the relevant chemical properties of crystals and how they interact with the environment lets us make a prediction of the form a crystal will take before we see it form. This is in fact routine for chemists in a lab, when a new salt is prepared, to simply assert that such-and-such is the form of the salt’s crystals. This prediction factually works as knowledge because the principles, or in this case – knowing the relevant material conditions on which the inference is to be made, are also knowledge.

There is a difficulty with the second claim being made specifically in the context of AlphaFold and similar AI systems. The issue essentially comes down to whether we can reasonably assume that AlphaFold has figured out some general principle or a mechanism of protein folding. I am wary of suggesting that there has to be a *law* of protein folding to solve this problem, but it is reasonable to posit that there has to be a *mechanism* of how proteins fold, meaning – if you know what the mechanism of amino acids interacting with other molecules is, then you can infer the shape of a protein based on this knowledge. In fact, the only way to fully hold the propositions (1) and (2) of the trilemma is to say that AlphaFold has implicitly grasped a theory of protein folding, and that this theory is scientific knowledge. Before I can discuss whether the theory can count as scientific knowledge part of this claim, I have to address the AlphaFold learning the theory.

In examining the reasons to think there is a general principle of protein folding, the two papers by Dill and colleagues (2012, 2008) provide substantial insights. First, proteins fold due to specific physicochemical forces encoded in their amino acid sequences, suggesting a general principle or a mechanism underlying the folding process. The mechanism is thought to dictate how proteins achieve their stable, functional native structures from their linear sequences, in the extremely short time they do so in nature. Finding out the relationship between a protein's sequence and its structure is a cornerstone of the folding principle. It is suggested that the amino acid sequence inherently contains the information needed to achieve the correct 3D shape. For more detailed discussion of the modelling suggestive of the possible protein folding principles and further evidence, see Dill (2012, 2008). When AlphaFold's creators at DeepMind claim it has solved the protein folding problem for science – we can assume they also believe it has figured out a robust mechanism.

One potentially relevant question, which I however will not pursue in this paper, is the nature of the mechanism of protein folding in nature and how it metaphysically relates to whatever it is a transformer AI architecture may be doing. This question concerns the nature of mechanisms and the role of theory in biology and their relation to laws in other sciences to which this problem is connected – physics and chemistry. Whether AlphaFold's solution represents the 'real' mechanism or there may be something else in nature that is essential to the potential theory of protein folding is an issue for a separate investigation. The question I do aim to get into is about how what AlphaFold is doing relates to how scientific knowledge is generated. Consider the following:

- AlphaFold was trained on everything we know about proteins and their structures so far.
- Based on all this data – it produced novel inferences (a massive amount of new data) that turned out to be highly reliable.

Potentially with AlphaFold, we have to think of it not as an *artificial scientist that discovers the mechanism* of protein folding by making tweaks to the theory but as an entity that embodies *a simulated principle, or model* of protein folding, which the scientist is tweaking in order to make this discovery. Consider that AlphaFold, just like many other AI deep learning systems, is trained on a set of data and uses some form of supervised and unsupervised learning. An essential step in training is when error of the output is minimized by changing parameters until the output matches what we already know to be true (feedback). The system is given a set of amino acid strings for proteins whose shapes we already know and then its parameters would be tweaked until it produces a correct output. One could point out that it is precisely a feature of AlphaFold, that even during this training step, tweaking the system's parameters does not give us an understanding of the mechanism, i.e. we do not gain knowledge of how and why the transformer architecture ends up distributing weights, what it picks up on, etc. It is almost like we are tweaking the numbers blindly, until it works. Let us now compare this with how a scientist adjusts parameters and calculations in a specific theory, calibrating it e.g. for making better predictions. Something similar often happens when a scientist is working on solving a new problem or working to discover a principle, where she tries out various ‘tweaks’ to make it work. While she has reasons to try out the tweaks she does, she is still often exploratorily *trying things out* because she does not have the answer yet, she does not know the principle she is searching for that would yield successful novel predictions, until she does something right and thus discovers the principle. In this process of calibration, which in ‘normal science’ belongs to the development and strengthening of scientific theories and methods and in machine learning belongs to the training, there seems to be a certain analogy in so far as tweaking the parameters goes. An important difference is of course that a scientist refining a theory for better predictions can explain her reasoning for paying attention to particular things, choosing the adjustments she did, etc., at least to an extent, whereas a deep neural network cannot explain its reasons for assigning weights across statistical distributions.

Another potential analogy can be drawn between AlphaFold as a predictive model of protein folding and how in certain scientific domains, the predictive success of a model can lead to the acceptance of the model itself as a discovery. For example, quantum mechanics' ability to predict a wide range of phenomena with high accuracy has led to its broad acceptance and the discovery of new physical principles, even before some aspects were empirically confirmed. Just like the body of theory in this case is tested by its ability to make good predictions – so is AlphaFold’s.

All in all, considering that AlphaFold is producing novel highly accurate data based on all the previously available evidence on how proteins fold, and that we have good reasons to believe there is a general principle of protein folding to begin with, it looks like what AlphaFold is doing resembles normal science, occurring at an accelerated pace. Except, whatever it has learned in terms of general principles or if it ‘became one’ itself – is not known by *any* human. *That* is what makes the picture strange, rather than the process of how it learns to predict the protein shapes. This opacity is the main cause for concern with accepting both propositions (1) and (2). In the next section I discuss proposition (3) – the problem of opaque scientific knowledge.

1.3. Rejecting 3: Why scientific knowledge *can* be strongly opaque

This claim concerns the most complex part of the trilemma and the heart of the dispute with my empiricist sceptic. In this section, I will be granting the assumption that AlphaFold does indeed have an implicit general principle of protein folding, in order to focus on the question of whether a strongly opaque principle or model can be scientific knowledge.

The notion of opacity in generative AI systems is not always straightforward, as there is some controversy in whether the knowledge we do have of the architecture of the neural network and the inputs of the training data (which arguably involves a lot of theory-ladenness about proteins) would still plausibly allow for the system to be called strongly opaque. To this, I would clarify that the strong opacity in question is related only to the knowledge of the assumed principle of protein-folding, which I here grant as picked out by AlphaFold to generate successful predictions on a massive scale for such a complex problem. Since it is still fair to posit that human scientists do not know how proteins fold in nature, and we still have a hypothesis that there is some general principle by which they do – it is fair to posit strong opacity as characteristic of AlphaFold's epistemology, at least as of today.

AlphaFold, being at least at present a black box, is a candidate for producing strongly opaque knowledge, and I argued in the previous sections that both the output (predictions that are individually not empirically confirmed) and the principle or an underlying theory (which AlphaFold embodies or has figured out) are genuine candidates for scientific knowledge. I argued that everything about the predictions and the predicting system constituting knowledge are in itself sufficiently aligned with 'normal' scientific practice, only arguing against empiricism in regard to the nature of confirmation and relationship between prediction and theory / mechanism for knowledge in science. However, what does *not* align with the normal scientific practice – is the strong opacity featuring in the AlphaFold case and is by extension likely to characterize all science driven by similar AI systems.

The question behind the claim (3) could be formulated as following: Are the general principles /mechanisms of protein structure AlphaFold has implicitly grasped (but that no one knows how to decode) part of scientific knowledge? An empiricist might defend (3), *scientific knowledge cannot be strongly opaque.* Despite the ubiquitous use of AlphaFold's predictions, an empiricist still has a good reason to deny that the principle behind them is knowledge, since she would be tied to the absence of a mental state that is 'knowing' how proteins fold in any of the scientists working with AlphaFold's outputs. The tension at this point focuses on the accessibility requirement for knowledge, which relates to an internalist and thus an individualist picture of conditions for knowledge vs an externalist view that prioritizes the trustworthiness (and may apply to both individualist and collective knowledge).

In this section I argue for the following: While an internalist picture is insufficient for scientific knowledge in general, an externalist reliabilist picture can also fall short of accommodating the AI-driven scientific practice, unless paired with a functionalist framework which fully rejects the requirement of mental states and accessibility in generating scientific knowledge. I briefly outline the elaborate arguments others have developed on the inability of internalist individualist epistemology to adequately accommodate scientific knowledge, largely due to the very nature of the modern scientific practice (Hardwig 2010, Bird 2010; 2022).

Around the time of the rise of the current transformer systems (Vaswani et al. 2017), a discussion arose around the epistemic justification of their outputs (Duran and Formanek 2018). Duran and Formanek (2018) developed a computational reliabilism framework, grounded in process reliabilism, arguing that an agent's belief in a proposition derived from a computer simulation is justified if the simulation is a reliable process that consistently produces true beliefs. Establishing the reliability / trustworthiness of an opaque process itself is the first step in looking for a justification of AlphaFold's integration into the production of scientific knowledge. However, there is more to the argument against (3) than confirming that AlphaFold's predictions are highly reliable and can therefore be taken as scientific claims. It is also about whether the part that remains opaque – the internal process that simulates protein folding – can be part of scientific knowledge. This is both connected to the justification of the individual predictions as scientific claims and calls for its own justification as part of knowledge, since we do not want to land in a situation where science is done by a process not known to science. Where we do eventually want to land is science being done by a process *not known to individual scientists* but *known to science as an enterprise*.

Alexander Bird's (2010; 2022) account of scientific knowledge as irreducibly *social / collective* knowledge can help accommodate the problem. The general (relevant) features of his view of science include:

- Science is a social enterprise, which can be explained as functionally analogous to but not supervenient on individual cognitive systems.
- Scientific knowledge is necessarily shared (or "social / collective") knowledge, as opposed to individual knowledge (does not include "belief" in definition)
- The argument stresses the division of labor in science and the complexity of the processes through which science is done ("doing normal science" (Bird 2010)) to reject individualist and internalist accounts of scientific knowledge.

To qualify for knowledge, Bird (2010) outlines three conditions. These can be sufficiently satisfied by AlphaFold:

1. *Outputs must be propositional in nature (propositionality)*. Both the predictions and arguably the underlying principle of protein folding (assuming there is one in this case) can be expressed in propositional terms and amount to scientific claims.
2. *Mechanisms whose function is to ensure or promote the chances that the outputs are true / valid / trustworthy are in place (truth-filtering).* AlphaFold is not an oracle or an alien artifact, but a learning system created by humans, which bases its novel outputs on all of our previous scientific data and claims. It is also integrated into the social system of science – there are peer reviewed papers on the quality of AlphaFold's predictions, the advancements in protein folding-related problems in biological sciences, on its use in drug discovery and design research, etc.; and it is integrated with various other processes and tools of science, which continuously and as a compound correct for errors and unreliable data.
3. *The outputs are the inputs for a) social actions or for b) social cognitive structures (incl. the very same structure) (function of outputs is preserved) – the outputs must be usable to produce more scientific knowledge.* AlphaFold's outputs are used in complex projects, as part of the division of labor between various other computational systems and human scientists. New cutting-edge research projects employ AlphaFold to varying degrees. Computational methods of problem-solving in science and in biological sciences specifically is advancing due to systems like AlphaFold, also a new version of the AI system is developed on the basis of the previous one[8], adapted to expand on the type of predictions and simulations it can produce.

The justification here is an externalist reliabilist one and is realized through the truth-filtering / trustworthiness condition. Since the functionalist account rejects the central role of individual minds in the knowledge production – the proposition does not need to be individually known by anyone, but it must be integrated into the broader scientific knowledge infrastructure, as the conditions 2 and 3 outline.

Further relevant features of his view lay the basis for decoupling of mental states from social knowledge, arguing that their role is peripheral, such that something can be knowledge without mental states involved:

- For P to be knowledge, no individual human has to know P at any given time (Bird 2010).

[8] The currently newest version is AlphaFold 3 (Abramson et al 2024).

- Though the process of P becoming knowledge usually, or at least so far, *involves* individual mental states (e.g. it is still largely humans who write and peer-review papers) but P does not become or remain knowledge *in virtue of being accessed* by individual mental states (e.g. the paper contains knowledge not in virtue of humans being able to read it but in virtue of it having been reviewed and utilized for producing further papers. If a machine could write and review a paper and if legitimate science could be done on the basis of it – it does not matter that humans have neither written nor accessed it).

Here arises a question of whether *accessibility* is a crucial part of the functionalist profile of knowledge for Bird's account, beside *trustworthiness* – being relied upon as a basis for further research, etc. That is, whether, even if no mental states are required to be present for something to continue to exist as part of scientific knowledge at any given time, does there have to be a mechanism in place allowing mental access to knowledge for the whole system to function properly? Bird claims explicitly that it is not access that defines the knowledge but the functional integration of knowledge into a societal structure: "It is not the accessibility of knowledge that is essential to being social knowledge; rather it is the capacity of the knowledge to play a social role (e.g. decision making by the group, a further research direction, etc.) in virtue of the structure and organization of the group; accessibility is the principal means by which that is achieved" (Bird 2010: 48). For him, however important the access to knowledge might factually be in current science, it is the function that plays the qualifying role. It is rather a way of instantiating a more general condition of social knowing, where accessibility comes in. In the AlphaFold case, one might note that since the implicit theory of protein-folding is not known by humans it cannot in fact play the social role, e.g. in decision-making. However, this is again an issue of conflating individual epistemic stance with the collective epistemological approach, as exemplified by Bird, where social role in decision-making need not be dependent on internal mental-state-like knowledge of individual scientists. E.g. if a question arises whether some drug modelled with the AlphaFold involvement should be developed further into an experimental or even trial phase, it does not come down to whether the predictions made by AlphaFold are transparent in their creation-mechanism. One might still object at this point, that, while the predictions may be accessed, the opaque mechanism cannot be used for social decision-making. At this point I see the distinction between the two to be quite thin. As far as AlphaFold can propose several structures to be taken as potential candidates for further research on a particular task – its internal mechanism is effectively driving the decision-making for the research direction, even if the human scientist downloading and further utilizing the proposed structures only directly interacts with the predictions in form of outputs and not with the internal mechanism itself.

However, one could imagine an objection to the functional mechanism of knowledge that excludes mental states completely. First, one could have a more general intuition that the scientific method still has to involve, to at least some degree, individual mental states in order to secure its legitimacy. That is because it promises, in theory, the *accessibility* to any individual that might want to 'check' a particular scientific finding, piece of data, etc. Following that intuition, one could argue that scientific knowledge can only be counted as such if it is technically accessible to any individual, independently of the societal structures and their functional mechanisms. Thus, one could claim that even though the functionalist account attempts to exclude cognitive states' role from the mechanisms of science as a collective enterprise, the role of the (mental states) access to scientific knowledge cannot be fully excluded from these mechanisms. The response to this part of the objection basically points out the conditions for knowledge defined by Bird and his motivation for the functionalist account – namely, that it is not actually possible to know most of the modern science in an individualist internalist way, if one seriously turns to mental states as the defining factor of knowledge (Bird 2010, 2022). A theoretical promise of cognitive access to knowledge simply does not suffice as a condition for scientific knowledge.

How does the AlphaFold case push this account towards a more radical version that is better able to accommodate AI-driven science? It at least seems to require a choice on whether trustworthiness without accessibility is enough, as AlphaFold may be satisfying the conditions for the former but not the latter. One could still insist that in an example such as the book sitting in a library and not one person in the whole world reading it, where the knowledge is qualified as such by being put into a book and the book in a library (both being instruments of knowledge in a socially defined mechanism) still implies that this mechanism is structured in the way that guarantees the access by means of a mental state to the knowledge kept in the book. However, I would argue that the fact that the access to this knowledge kept in the book is ensured by the features of the mechanisms we build (the book being readable, the libraries accessible, etc.) is due to the fact that individual humans have so far played a contingent, if factual, driving role in utilizing the knowledge and defining its trajectories. Being cognitive agents able to possess individual knowledge – we mostly used to build things cognitively accessible to us. At least we had done so for a long time, before the new AI systems have gradually started to challenge our cognitive capacities, without it necessarily being our direct intent[9]. That is to say, the fact that the book and other human artifacts are at least typically cognitively accessible to individuals ultimately only matters contingently in how the knowledge they keep performs its (societal) function.

The decoupling of scientific practice and mental states thus can be expanded on with AI-driven science, where e.g., AlphaFold produces outputs which satisfy the externalist functionalist conditions to qualify as knowledge, on the view where mental states are strongly excluded. Another choice AlphaFold and similar systems push us towards is between trustworthiness and accessibility. While accessibility is central for trustworthiness on an internalist view, it is not necessary for trustworthiness on an externalist view (letting it come apart), and it is arguably only tangential to scientific knowledge overall on the functionalist externalist view of collective social knowledge.

I have defended the claim that scientific knowledge can be strongly opaque by arguing that:

- The functionalist account of scientific knowledge allows for knowledge to be strongly opaque (even if it is normally not), if it is otherwise reliable and properly functionally integrated into a collective scientific enterprise.
- AlphaFold is (or can be) properly functionally integrated into a collective scientific enterprise.
- AlphaFold generates scientific knowledge, part of which is strongly opaque.

The outlook on the wide employment of systems such as AlphaFold in current and future science generally confronts us with the genuine possibility that strong opacity of the mechanism is largely compatible with the rich functional integration with the rest of science. If the integration is in place and AlphaFold functions as a part of a larger system that sustains itself according to the scientific norms and standards – the inaccessibility of its internal principle should not prevent us from treating it as scientific knowledge.

2. Objections

I argued that a version of an externalist functionalist account of knowledge can successfully accommodate the AI-driven scientific knowledge production if one fully rejects accessibility

[9] Here I refer specifically to the opacity of the AI-systems not being intended as such but constituting the result of the specific ML methods being applied in a broader framework – a race towards creating artificial intelligence and the most powerful artificial systems.

to knowledge as a required feature of its functionalist profile. Still, there may be a skepticism about externalist reliabilism properly accommodating scientific knowledge that is opaque.

Such skepticism may be motivated by the famous case studies from reliabilism, such as the "chicken-sexers" case[10]. Chicken-sexers are individuals that are uniquely skilled in determining the sex of chickens on sight. Intriguingly, as the case is reported, they often cannot articulate the criteria they use in how they actually sort the chicks. The reliabilist externalist claim in this case is that as they are highly reliable in sorting chickens, their propositional belief that "this is a male chick" or "this is a female chick", etc. are justified, even though they lack an internalist justification – cognitive access to reasons for their own judgements. Their judgments about a chick's sex are considered knowledge on this account, even though they lack the ability to provide an explicit, internally accessible justification for their decisions. Proponents of internalism typically respond to this scenario in one of two ways: either by contesting the notion that chicken-sexers possess genuine knowledge, or by arguing that there must be subtle, perhaps subconscious, cues in the chicken-sexer's perceptual experience that guide their determinations. For an empiricist – accessibility is an important feature she would not easily let go of. While one may plausibly defend that individual chicken-sexers may possess genuine *individual knowledge*, it is hard to imagine a *scientific paper* claiming that a chicken-sexer has been used to determine the biological sex of chickens, but we do not have any idea how and what makes them reliable.

Another parallel a sceptic might draw is the "Truetemp" thought experiment (Lehrer 1990), which is commonly used to argue against (opaque) reliabilism (Goldman 1994). According to the thought experiment, Mr. Truetemp is magically and unbeknownst to him given an unusual but highly reliable cognitive faculty of measuring the exact temperature of the air. Since he is unaware of the existence of this faculty, its deliverances (suddenly "knowing" the temperature and always finding out he is correct about it) strike him as rather odd. Many internalists think that Truetemp would not have justified beliefs. Following this intuitive judgement, since he clearly satisfies the reliabilist conditions for justified belief – reliabilism appears to be mistaken. The defense of Truetemp's beliefs as genuine knowledge has also been focused on determining the existence of some "hidden" or "underlying" (potentially meaning 'unconscious') cognitive access to his temperature measuring mechanisms, that would allow for valid justification. Beebe (2004) for example, argues that the case is under-described, that

[10] https://iep.utm.edu/int-ext/

the new cognitive faculties need time for the individual to adjust to, which would arguably lead to a form of cognitive access to the belief and then the belief would be justified.

Cases like chicken-sexers and Truetemp can motivate an empiricist sceptic, claiming that it is resembling AlphaFold in important respect – the opaqueness of the mechanisms that make the propositional output reliable. Just as those cases cannot constitute scientific knowledge – what the AlphaFold is doing cannot either.

To respond to this objection, I want to bring the focus back onto the functionalist view of scientific knowledge as social or collective knowledge, which is necessarily such and cannot be reduced to individual knowledge. I am framing the AlphaFold epistemology as a case not for reliabilism per se (in its individualist form) but rather for radical externalist functionalism; the key here is integration of the given knowledge candidate into science as a system and rejecting the core dependence on individual knowledge in science. That means, that the opacity to individual scientists is itself not the issue. However, the issue is whether there is opacity to the system of science as a whole. AlphaFold is integrated into the system of science, because it is operating based on what is known to science independently of AlphaFold (its training data, machine learning), its training is finetuned to produce high-accuracy outputs, and the outputs themselves have been proven to be reliable and are ubiquitously used in further scientific research. The conditions Bird (2010) outlined for mechanisms whose function is to ensure or promote the chances that the outputs are trustworthy to be in place, and for the outputs to function as further inputs for the scientific collective knowledge and the enterprise of science are thus sufficiently satisfied.

Thus, the general response to the objections from internalists about reliabilism being problematic for scientific knowledge is that the reliabilist externalism which focuses on the individualist knowledge conditions is just as unable to accommodate scientific knowledge as internalism. AlphaFold should not be compared to an individual scientist / person, whose mind is opaque to other people, but rather understood as a system that is part of science as a collective enterprise. That part of the scientific enterprise may be opaque to people, but its processes and mechanisms are made trustworthy by the social enterprise and mechanisms of science.

Conclusion

I have presented the epistemology of AlphaFold as a trilemma and argued against a potential empiricist sceptic that scientific knowledge can be opaque to humans, as long as it is properly functionally integrated into the scientific practice. If it is sufficiently integrated – it allows us to ground the claim that AlphaFold generates scientific knowledge. In accommodating such systems epistemologically, we come closer to developing epistemology that is more responsive and appropriate to the current practices in the AI-driven science.

If we accept a version of an externalist functionalist view which completely rejects the role of accessibility to knowledge through mental / cognitive states, we can argue that AlphaFold generates scientific knowledge, even if a part of that knowledge is strongly opaque. The tension remains between a more 'traditional' internalist empiricism and an externalist realism. Given that AlphaFold's predictions are taken to be trustworthy and used by scientists to conduct further research before empirical confirmation pushes us to broaden the notion of the relation a prediction needs to stand in to scientific knowledge. Otherwise, we find ourselves approximately in a situation where scientists are advancing science without knowledge. The AlphaFold case in a way forces us to take a position on whether we can call something that is reliable, novel, based on existing scientific knowledge, and used ubiquitously for cutting-edge science – scientific knowledge, if it is not directly empirically confirmed and may never be individually known to humans. Possible implications of such AI-driven scientific practice include a picture where what can be plausibly considered scientific knowledge may come further apart from understanding and explanation as the goals of science.

Bibliography

Abramson, J., Adler, J., Dunger, J. *et al.* 2024: Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

Al-Janabi, A. 2022: Has DeepMind's AlphaFold Solved the Protein Folding Problem? *BioTechniques*, *72*(3), 73–76. https://doi.org/10.2144/btn-2022-0007.

Beebe, JR., 2004. Reliabilism, Truetemp and New Perceptual Faculties. *Synthese* 140: 307–329.

Bird, Alexander 2010. 'Social Knowing: The Social Sense of "Scientific Knowledge": Social Knowing'. Philosophical Perspectives 24 (1): 23–56. https://doi.org/10.1111/j.1520-8583.2010.00184.x.

Bird, A., 2022. Knowing Science. First edition. New York: Oxford University Press.

Boyd, NM., Bogen, J., 2021: "Theory and Observation in Science", *The Stanford Encyclopedia of Philosophy,* Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2021/entries/science-theory-observation/>.

Dellsén, F., 2016: Scientific progress: Knowledge versus understanding. *Studies in History and Philosophy of Science*, 56, 72–83.

Dill, KA, MacCallum, JL., 2012: The Protein-Folding Problem, 50 Years On.*Science***338**,1042-1046. DOI:10.1126/science.1219021.

Dill, KA, Ozkan SB, Shell MS, Weikl TR., 2008: The protein folding problem. Annu Rev Biophys;37:289-316. doi: 10.1146/annurev.biophys.37.092707.153558. PMID: 18573083; PMCID: PMC2443096.

Durán, J.M., Formanek, N., 2018: Grounds for Trust: Essential Epistemic Opacity and Computational Reliabilism. *Minds & Machines* **28**, 645–666. https://doi.org/10.1007/s11023-018-9481-6.

EMBL-EBI 2022: DeepMind AI Powers Major Scientific Breakthrough: AlphaFold Generates 3D View of the Protein Universe. SciTechDaily https://scitechdaily.com/deepmind-ai-powers-major-scientific-breakthrough-alphafold-generates-3d-view-of-the-protein-universe/.

Goldman, AI., 1976: "Discrimination and Perceptual Knowledge", *The Journal of Philosophy*, 73(20): 771–791. doi:10.2307/2025679.

Goldman, AI., 1986: *Epistemology and Cognition*, Cambridge, MA: Harvard University

Press.

Goldman, AI., 1994: Naturalistic Epistemology and Reliabilism. Midwest Studies In Philosophy, 19: 301-320. https://doi.org/10.1111/j.1475-4975.1994.tb00291.x.

Hardwig, J., 1985: Epistemic dependence. *Journal of Philosophy*, 82, 335–49.

Hardwig, J., 1991: The role of trust in knowledge. *Journal of Philosophy*, 88, 693–708.

Harker, D., 2008: On the predilections for predictions. *British Journal for the Philosophy of Science*, 59, 429–53.

Heikkilä, M., 2022: DeepMind has predicted the structure of almost every protein known to science. MIT Technology Review.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, L., et al. 2021: 'Highly Accurate Protein Structure Prediction with AlphaFold'. Nature 596 (7873): 583–89. https://doi.org/10.1038/s41586-021-03819-2.

Shapere, D., 1982: The concept of observation in science and philosophy. Philosophy of Science 49 (4):485-525.

Lehrer, K., 2000: Theory of knowledge. Boulder, Colo.: Westview Press. MIT Technology Review:https://www.technologyreview.com/2024/10/09/1105335/google-deepmind wins-joint-nobel-prize-in-chemistry-for-protein-prediction-ai/.

Norton, JD., 2021: The Material Theory of Induction. University of Calgary Press.

O'Connell, L., 2018: The Meckel–Serres conception of recapitulation. *Embryo Project Encyclopedia*. accessed 1 March 2021.

Poston, T: Internalism and Externalism in Epistemology. Internet Encyclopaedia of Philosophy. https://iep.utm.edu/int-ext/ (accessed 05.11.2024).

Press release. NobelPrize.org. Nobel Prize Outreach AB 2024. Tue. 5 Nov 2024. https://www.nobelprize.org/prizes/chemistry/2024/press-release/.

Quach, K., 2022: Nearly all protein structures known to science predicted by AlphaFold AI. Let's hold off the champagne until an actual drug is developed using this tech. Enterprise Technology News and Analysis.

Ren, F., et al., 2023: AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. Chem. Sci., 2023, 14, 1443. DOI: 10.1039/d2sc05709c.

van Fraassen, B., 1980: *The Scientific Image*, Oxford: Oxford University Press.

van Fraassen, B., 1985: "Empiricism in the Philosophy of Science", in Churchland and Hooker 1985, pp. 245–308.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, AN., Kaiser, L., Polosukhin I., et al. 2017: Attention is all you need. https://doi.org/10.48550/arXiv.1706.03762.

Yang, Z., Zeng, X., Zhao, Y. et al., 2023: AlphaFold2 and its applications in the fields of biology and medicine. *Sig Transduct Target Ther* **8**, 115. https://doi.org/10.1038/s41392-023-01381-z.