Structure Representation of Deep-Learning Models: The Case of AlphaFold

Giovanni Galli

University of Urbino Carlo Bo

Abstract

The scientific enterprise enriches the debate about models. In particular, in the field of structural biology, a new deep-learning neural network system called AlphaFold has been applied for many purposes. It allows us to predict a protein's structure with high accuracy. I will present the system in light of the discussion of structure representation and argue for a specific kind of representational relation holding between the predicted model structure and its target-system. By doing so, I will criticize the artifactual approach advanced by Knuuttila (2021) and present the features that characterize the predicted structures of AlphaFold as simulation models.

Keywords: Scientific representation, Deep-learning models, AlphaFold, Protein structure determination.

1. Introduction

The notion of model is one with a wide polysemy within the sciences and philosophy. There is no unique conceptual framework and definition able to define all the models involved in scientific activities. There is no broad consensus on any unified account of models, as stated by Gelfert (2017), and it is considered an obvious consequence of this void to assume that "if all scientific models have something in common, this is not their nature but their function" (Contessa 2010: 194). Moreover, if this characterization of models as functional entities is accepted, we must then specify how the models work as "carriers of scientific knowledge" (Ducheyne 2008: 120).

One of the basic relationships between the model and its target-system (T) that has to hold, if the model must carry scientific knowledge, is the representation.[1] My aim is not to advance a general theory of scientific representation, but

[1] See also Campbell 1920; Hesse 1966; Giere 1988; Morgan and Morrison 1999; Hughes 1997; Teller 2001; Van Fraassen 2008; and Mitchell 2013. Concerning the issue of scientific representations and realism, deeply tight, for a defense of scientific realism, see also Alai 2021a, 2021b, and 2023.

Argumenta 9, 1 (2023): 43-60
ISSN 2465-2334

DOI 10.14275/2465-2334/202317.gal

to propose a definition of the representational relationship between the specific kind of models produced by the deep-learning neural network system AlphaFold (AF), and their T. In §2 I present the main positions and definitions of models as functional entities. This, however, is mostly a study about the semantics of the representational relationship between AF and its T, for it is on the basis of that relation that such models are carriers of knowledge. Examples of this relationship regard models of actual T, such as the double-helix model of DNA, or the Bohr model of the atom, i.e. models that represent existing objects, and also models of potential (non-actual) T, as the examples of repressillators, synthetic oscillators and the ultra-Keynesian model analyzed by Knuuttila (2021), i.e. models that represent objects not existing in nature. According to the *representationalist* view what we learn from models presupposes a representational relation, while according to the *inferentialist* view, the representational feature of models is decoupled from their capacity of carrying knowledge. I claim that the representational relation presupposes the epistemic function of models of both actual and potential T. In §3 I discuss Knuuttila's (2021) artefactual view of models. In §4 I argue that the example of models of potential T does not invalidate the role of the representational relationship, and in §5 I discuss the contest of Critical Assessment of protein Structure Prediction (CASP) and AF. In §6 AlphaFold models are interpreted as simulation models. To conclude, in (§7) I argue that they hold a kind of morphic representational relation with their T. The general aim of this paper is to give one of the first contributions to expand a philosophical account of deep-learning models in general and AF models in particular.

2. A Taxonomy of Models

Models have a central role in sciences. Even if there is no consensus about their nature and qualifications, scholars have elaborated on three main areas: semantics, ontology, and epistemology of models. The first relates to what the models represent. The second concerns what the models are. The third focuses on the cognitive function modelers exploit for epistemological purposes. I will focus mainly on the first area, addressing namely the relation between the model and its target—system, specifically in the context of material, artifactual and simulation models, as they are tackled by Rosenblueth and Wiener (1945), Knuuttila (2021) and Durán (2018, 2020).

There are three main conceptions of the model–T relation: the similarity conception, i.e., models and their T are to some extent similar; the structuralist conception, i.e., models represent their T in virtue of a morphic relation between them; and the inferential conception, i.e., models as scientific representations have to be analyzed in terms of the inferential function.[2] Each conception offers different answers to certain problems. Moreover, we can distinguish the instantial view and the representational view. According to the former, models instantiate the axioms of a theory, that is composed of linguistic and mathematical statements. The representational view instead holds that it is rather the language that is connected with the model, while the model connects to the world "by way of similarity between a model and designated parts of the world" (Giere 1999: 56). In turn, the representational view has an informational and a pragmatic version.

[2] For a general discussion about the arguments and problems of the three accounts of scientific representations, see Frigg and Nguyen 2021.

The former conceives representation as "an objective relation between the model and its target, which imbues the former with information about the latter" (Gelfert 2017: 26). According to the latter, instead, it is not possible to "reduce the essentially intentional judgments of representation-users to facts about the source and target object or systems and their properties" (Suarez 2004: 768).

A further distinction can be drawn between substantive and deflationary accounts of representation. Substantive accounts aim for a robust explanation of the function of a representation in terms of a fundamental relation between a model and its target. Deflationary accounts, instead, settle for a light characterization of the functional unit of representational devices. We will see that while Knuuttila's proposal is pragmatic and deflationary, even though recognizes a representational function of models, the AF models are better interpreted by the representational, informational, and substantive view.

3. The Artifactual Account of Models

AF models, as representations of proteins, are a result of sophisticated techniques that make use of experimental data and abstract models. The 3d structures of proteins predicted by AF recall the structure of material models of a DNA strand but with a digital suit. One of the first studies on the representational capacity of models has been made by Wiener and Rosenblueth (1945). They analyze the role of material models of phenomena in scientific research, stressing their advantage with respect to abstract models thanks to their representational features. They describe a material model as "the representation of a complex system by a system which is assumed simpler and which is also assumed to have some properties similar to those selected for study in the original complex system" (Rosenblueth and Wiener 1945: 317). The relation identified by the authors between the material model and the original complex system can be seen as a case of similarity conception. This view then contrasts Suarez's inferential conception. These models are intended to be approximations and "surrogates" (Rosenblueth and Wiener 1945: 320) for the real facts under observation. But models can represent also facts not already present in reality. Indeed, Knuuttila is interested in developing an account of models consistent with the need, in some areas of inquiry as economics or synthetic biology, to build models of objects we do not find in nature or in society, i.e. models of invented objects.

Knuuttila (2021) advances the artifactual account of models which fits well with the inferential account developed by Suárez (2004). She is interested in stating an alternative position to the received ones, both substantive and deflationary, pointing out that models can be carriers of scientific knowledge even if they do not represent the actual state of affairs in the world. She insists on the modal reach (Godfrey-Smith 2006) and the modal dimension of modeling (Le Bihan 2016), "which approaches models as purposefully constructed systems of interdependencies designed to answer some pending scientific questions" (Knuuttila 2021: 5). Models as epistemic artifacts function as "erotetic devices" (Knuuttila 2021: 6). Such devices are artificial systems that deploy dependencies constrained to the aim of answering a specific scientific question, supported by theoretical, and empirical considerations.

Two examples are described, one of an ultra-Keynesian model as an example of an economic model that does not refer to a real T, and one of repressilators and synthetic oscillators in synthetic biology, that do not correspond to any existing

circuits, but are rather pictured to explore and test possible biological circuit designs. To strengthen the cases, she distinguishes between representational modes and media, and also between internal and external representations. The representational modes are the many semiotic devices that express various meanings and contents, while the representational media are for example the ink on paper, digital computer, biological substrata, and what support the representations. According to Knuuttila (2021: 5) the same representational mode can be implemented in different media as the example of the synthetic repressilator and the electronic repressilator that instantiate both the same ring oscillator design, yet they are implemented in different media "enabling different kinds of inferences" (Knuuttila 2021: 5). Moreover, an internal representation concerns "how various kinds of sign-vehicles or representational devices are used to make meaning and convey content" (Knuuttila 2021: 5), i.e. for a material model of the atom, the material, the proportion, and in general the semiotic and semantic features of the model chosen to represent the specific object; by external representation, instead, she refers "to the relationship of a model to a real-world target system, the question on which the philosophical discussion has largely concentrated" (Knuuttila 2021: 6). This distinction is particularly relevant for the definition of models as epistemic artifacts:

> Nevertheless, the fact that something may be internally represented within a model without necessarily representing the actual state of worldly affairs opens up the prospect of conceiving modeling as a practice of exploring the possible (Knuuttila 2021: 7).

The artifactual approach allows us to see biology as a discipline that not only focuses on natural organisms but includes also potential organisms (Elowitz and Lim, 2010, 889). So conceived, models are carriers of knowledge in virtue of their being erotetic devices and artifactual constructs useful to support surrogative inferences about a potential target-system. In such a way, inferentialists would argue that their representational capacity is not relevant to their use in exploring the possible.

4. Some Remarks on the Artifactual Account of Models

The artifactual account stresses the pragmatic goal that directs the models' construction and manipulation. It is to conceive models as tools for investigating specific phenomena, used to answer scientific questions, motivated by theoretical, and empirical tenets. According to Knuuttila (2021), their accomplishment relies on their modal function of exploring the spaces of possibilities and the main point is that their success needs not be grounded on the representational relation between the model and the target system. Thanks to the distinction between internal and external representations, Knuuttila safeguards a slightly deflationary definition of representation, which connects the artifactual models with a possible organism. Obviously, the correctness of models of merely possible T does not need the same kind of warrants as the models of real T. What does then warrant them? For Knuuttila it is simply their predictive success, without any need to invoke to any representational relation, yet it remains unanswered the question concerning what warrants the models' success. In other words, how can we probe the success

of a model of a potential target-system, without any reference to the representational relation between the model and the possible state of affairs? Knuuttila (2021) claims that it is still sufficient for a modal relation to justify the success of the artifactual models.

We can reframe the modal feature of the relation between the models and the potential T as a predictive relation, i.e., a model would predict the possible state of affairs, if there were conditions such and such. One of the kinds of models so far used to explore the possible phenomena within a manifold scenario is the simulation model (SM). That is a model resulting from computational procedures able to predict or determine specific output with a given set of data. SM are helpful to study and predict complex scenarios and phenomena. They are implemented by a certain degree of idealization and can be used to study *actual* T (like biological systems, i.e. birds flocks, ant colonies, structure determination, enzyme kinetics and molecular dynamics) and *potential* T (like the behavior of mechanics and artifacts as airplanes, spacecrafts, biomedical robots, and also new proteins, new drugs and possible organisms). As it happens with imaginary economics, repressilators and oscillators, from a set of data and techniques the respective models predict how the possible systems would act. To this extent, artifactual models are a kind of simulation model: though the examples are not strictly speaking computer-based simulations, they simulate possible states of affairs, useful to predict how the system will work.

I submit, however, that neither for simulation models nor for material models we can easily dismiss the representational link between the model and its T. In the case of artifactual models, it seems intuitive not to stress the representational link, because we weigh differently the conceptual role of an actual T and a potential T. However, if we want to gain epistemic access to the T in question, actual or potential, the model has to maintain a representational link with it. I call it the accessibility condition (AC):

> *Accessibility condition*: A model M of a target-system T is a functional carrier of knowledge in virtue of its capacity to give epistemic access to T through the representational relation established by the researchers between M and T.

In the case of AF the output models of predicted proteins' structures can be conceived as a kind of artefactual model. Most AF models represent actual target systems, but they are also useful in the exploration of potential proteins. In that case, their success depends on their accurately representing the modal properties of proteins, i.e., what is actually possible or impossible for proteins. The discussion on representation, then, is far from over, and a substantive view of representation is still in play.

5. CASP and AlphaFold Protein Structure Prediction

AF is a breakthrough deep-learning network AI system able to predict highly accurate protein structures.[3] Its computational power and sophisticated engineering let the DeepMind team, which worked on it, win the CASP 14 (Critical Assessment of Protein Structure Prediction) on the 30th of November 2020. CASP started in 1994 and it is a biennial competitive appointment for biological researchers working on protein structure prediction, aiming at solving the well-

[3] All the predicted structure can be found on the AF open access database here: https://alphafold.ebi.ac.uk/ (last access November 2023).

known folding problem: How is it possible to fold a protein starting from its strains of amino acids? The founder and chair of CASP is John Moult, Professor of the Institute for Bioscience and Biotechnology Research and the Department of Cell Biology and Molecular Genetics at the University of Maryland. He describes CASP in this way:

> Computational biology differs from traditional science in that it takes place in a virtual world. Achieving rigor in a computational world which the scientist controls is much harder than when dealing with the inflexible realities of the physical world. We introduced Community assessment experiments in computational biology to help achieve the same rigor as in real world science. CASP (Critical Assessment of Structure Prediction), the first framework for these experiments, is an organization that conducts double blind community wide experiments to determine the state of the art of computational methods for modeling protein structure from amino acid sequence and other information. CASP has now been running for over 20 years, with continuing high participation rates (over 100 groups around the world), and has been accompanied by an enormous improvement in the accuracy of the protein modeling methods. The CASP methodology has now been adopted in a wide range of computational biology areas, including protein-protein interactions, genome sequence annotation, biological networks, and protein function annotation (Moult 2022).

The first lines make a sharp distinction between the rigor achieved in the real-world sciences and the one obtained in a computational world. I am interested in showing the philosophical relevance of the effort to make the two methodologies meet and enhance each other. Two questions. Why do the real-world sciences working on protein folding need such an upgrade? Moreover, why is it so important to solve the folding problem? "We have discovered more about the world than any other civilization before us. But we have been stuck on this one problem. How the proteins fold up. How the protein goes from a string of amino acids to a compact shape that acts as a machine and drives life?",[4] says John Moult (2021), filmed in *AlphaFold: The making of a scientific breakthrough*, the inside story of DeepMind[5] research team who created AF. This is indeed the folding problem. Solving it means making huge steps in molecular biology and consequently in many other biological fields. DeepMind team states that the research program that leads to AF and similar systems is crucial for the development of the life sciences. Proteins are stunning biological nano-machines, whose understanding will take us to unveil how they work and interact with other molecules. They are polymers in which the 20 natural amino acids are connected by amino bonds. They are polymers in which the 20 natural amino acids are connected by amino bonds. They are synthesized by the ribosomes, which are complex molecular machines present in all living cells, measuring around 30 nm. Ribosomes compose amino acids together in the specific order defined by messenger RNA molecules.

[4] John Moult was interviewed in *AlphaFold: The making of a scientific breakthrough*, video interview about the AlphaFold breakthrough: https://www.youtube.com/watch?v=gg7WjuFs8F4 (last access November 2023).

[5] AlphaFold thematic section on DeepMind website: https://www.deepmind.com/research/highlighted-research/alphafold (last access November 2023).

AF team trained this system[6] on publicly available data consisting of around 170.000 protein structures taken from the protein data bank (PDB),[7] together with large databases containing protein sequences of unknown structure. Thanks to the genomics revolution we can read amino acid sequences of proteins at massive scale; in fact, the Universal Protein database (UniProt) contains 180 million protein sequences. The building blocks of proteins are amino acids, small molecular compounds with unique features composed of between 10 and 20 atoms. In ordinary biology we find 20 standard types of amino acids floating within the cytoplasm of the cells. They connect to a piece of transfer RNA that matches with the three genetic sequences of the genetic code of the RNA messenger. Ribosomes then read the three-basis instructions of the RNA messenger and start building a chain of amino acids that goes out from the ribosome. As the chain of amino acids exits the ribosome, released in the cytoplasm, it is surrounded by water molecules and subject to the interaction of physical forces that make the chain fold up on itself and form the complex 3d structure we call a protein. All this process is called translation because the molecular mechanisms manage to produce a fully operative protein with proper functions by translating a piece of the genetic code. The unique shape of a protein is defined by its amino acid sequence and its shape is the key to unlock its functions. Determining the 3d structure of a protein is indeed necessary to understand its functions. Proteins seem like pieces of a puzzle, but with a dynamic shape which can change according to the bonds they make with other interacting molecules. Nonetheless, a protein would bond with some molecules and not others. There are specific combinations of proteins and molecules. By understanding the protein shape and the occurring molecular interactions, scientists can design vaccines, new drugs and functional structures for ecological purposes: "Among the undetermined proteins may be some with new and exciting functions and—just as a telescope helps us see deeper into the unknown universe—techniques like AlphaFold may help us find them" (The AlphaFold Team 2020).

Proteins are fundamental for most living beings, and enhancing their understanding through computational allows us to tackle diseases, discover new medicines and disclose the enigmas of life in a faster and cheaper way than traditional research on existing proteins. Thanks to painstaking experimental effort, real-world sciences have determined before the release of AF the 3d structures of approximately 100.000 unique proteins (Thompson, Yeates and Rodriguez 2020; Bai, McMullan and Scheres 2015; Jaskolski, Dauter and Wlodawer 2014; Wüthrich 2001). Using the experimental methodology scientists had at their disposal until now, it could take from months to years and a lot of financial resources to determine a single protein structure. Computational methodologies are in fact needed to reduce this gap and to "enable large-scale structural bioinformatics" (Jumper, Evans, Pritzel et al. 2021a: 1). That is why CASP has been promoted within the biological fields, with the aim to push researcher communities to solve the protein folding problem, that has been an open research problem since when,

[6] It makes use of 16 TPUv3s (which is 128 TPUv3 cores or roughly equivalent to ~100-200 GPUs) run over a few weeks, a relatively modest amount of compute in the context of most large state-of-the-art models used in machine learning today. See Jumper, Evans, Pritzel et al. 2021a.

[7] Protein Data Bank website: https://pdb101.rcsb.org/ (last access November 2023).

around 1960, the first atomic-resolution protein structures were proposed (Kendrew 1961; Pauling and Corey 1951; Pauling, Corey and Branson 1951), while the first protein structures detected presented unpredicted irregularities. It was the case of globin structures, a clade of globular proteins containing heme, a precursor to hemoglobin (6,5 nm), involved in binding, and transporting oxygen. Globin proteins contain the globin fold, which is a series of eight α-helices packed together in irregular ways. Since the 60's the folding problem concerns three different problems (Dill, Ozkan, Shell and Weikl 2008):

1) The folding code: the thermodynamic question of what balance of interatomic forces dictates the structure of the protein, for a given amino acid sequence;
2) Protein structure prediction: the computational problem of how to predict a protein's native structure from its amino acid sequence;
3) The folding process: the kinetics question of what routes or pathways some proteins use to fold so quickly. We focus here only on soluble proteins and not on fibrous or membrane proteins.

The main CASP evaluation follows the criteria of comparison between the predicted model α-carbon positions and those in the real-world target structure. The visualisation of cumulative plots of distances between pairs of α-carbon in the model and target structure positioning is used to evaluate the prediction against the experimental result, such as shown in the two figures aligning computational prediction with the experimental result. The real structure is already known by the evaluator so that the CASP examination can estimate the accuracy of the predictive model. To each prediction is assigned a numerical score GDT-TS (Global Distance Test—Total Score) specifying the percentage of modeling residues[8] in the model with respect to the target.

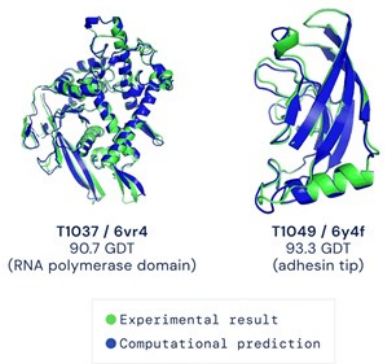

Figure 1: Two examples of protein targets in the free modelling category. AlphaFold predicts highly accurate structures measured against experimental result (The AlphaFold Team 2020).

The CASP campaign evaluation relies basically on the issues of 1) The folding code, 2) Protein structure prediction, and 3) The folding process, although the

[8] The amino acids in a polypeptide chain are linked by peptide bonds. Once linked in the protein chain, an individual amino acid is called a residue, and the linked series of carbon, nitrogen, and oxygen atoms are known as the main chain or protein backbone.

results are carried out in many prediction categories: tertiary structure prediction, residue-residue contact prediction, disordered regions prediction, function prediction, model quality assessment, model refinement, and high-accuracy template-based prediction. Tertiary structure prediction is then divided into three subcategories: homology modeling; fold recognition; and *de novo* structure prediction (New Fold). All these conditions form what we can call the *accuracy qualification* (AQ). The higher the GDT scores, the better the AQ of the predictions, and the higher the AQ, the nearer the model to the real shape of the protein. Another consequence of the AQ is that higher scores correspond to higher amounts of correct information transmitted from T to M, and from M to the modelers.

Since 2018 CASP team made some improvements, but the big leap was between AlphaFold 1 (AF1), the ancestor, and its successor, AlphaFold 2 (AF2), whose score, according to Moult, was around 90 GDT on 100 points scale prediction accuracy. DeepMind developed new deep learning architectures to improve the research methods for CASP14, which led to a high level of accuracy. These methods are inspired by the research areas of biology, physics, and machine learning and by the studies many scientists enhanced during the years on the protein folding problem. The AF2 system is described as a neural network-based model (Jumper, Evans, Pritzel et al. 2021a). It is important to note that it is described as an AI system coherent with the wider project of Demis Hassabis, CEO and co-founder of DeepMind, of making further steps in General AI. The whole AF architecture learns from the data and elaborates the 3d structure prediction of the folded protein. We can think of a folded protein as a spatial graph, a spatial presentation of a graph in the 3-dimensional Euclidean space R^3, in which residues are the nodes and edges link the closely related residues (Jumper, Evans, Pritzel et al. 2021a). The graph matters to understand the proteins physical interactions and their evolution. For the second version of AF2, the team created an attention-based neural network system, trained end-to-end, that attempts to interpret the structure of this graph while reasoning over the implicit graph that it's building (Jumper, Evans, Pritzel et al. 2021a). By process iteration, AF2 produces accurate predictions of the underlying physical structure of the protein in daystime. Moreover, the system can predict the reliability of parts of each predicted protein structure using an internal confidence measure. The following is the AF1 architecture that provided important results in CASP13, beating the median free-modeling accuracy of other systems.

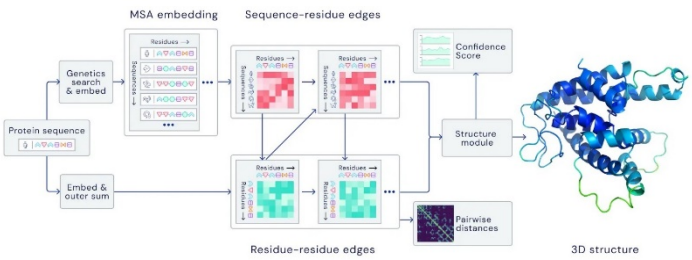

Fig. 2: An overview of the main neural network model architecture. The model operates over evolutionarily related protein sequences as well as amino acid residue pairs, iteratively passing information between both representations to generate a structure (The AlphaFold Team 2020).

AF1 has a straightforward architecture (Senior, Evans, Jumper et al. 2020). It begins with the amino acids sequence for which we are searching the protein structure. The first step concerns a data extraction move from the known database, in order to find similar protein sequences. The first task of the neural network is to find similar sequences, and it is called Multiple Sequence Alignment (MSA). The protein structure is responsible for its function, and we know that evolution carved the organisms in such a way that only some structures passed the survival threshold. Indeed, in different organisms during evolution a protein structure is more stable over time than the genetic sequence encoding that particular protein the genetic mutations that passed the evolutionary test are those that did not affect the protein structures. Comparing evolutionary-related protein sequences, whose 3d form should share some similarities, is what MSA does: scrolling the database to find amino acid sequence matches in the animal kingdom. To sum up, in AF1, 3 main steps need to aim at structure prediction:

1) AF1 collects the MSA features;
2) it predicts then the distogram using a residual neural-network;
3) it optimizes the protein backbone using the predicted distogram in combination with simulated physical forces. The output is the 3d predicted protein structure.

As the aforementioned system, AF2 presents three main blocks:

1) A pre-processing stage where the input sequence is used to query additional information about the initial sequence from databases;
2) The information is then mapped into an MSA and pair representation, which are refined by the Evoformer, a 48-layer deep transformer-like network that uses attention mechanisms to update MSA and pair representations;
3) The structure module, a recurrent network, processes the Evoformer output, which transforms the abstract representations of the Evoformer into concrete 3d coordinates of the protein geometry.

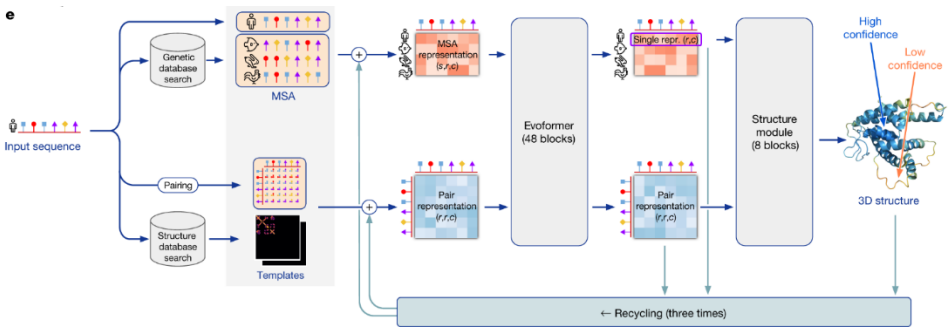

Fig. 3: Model architecture. Arrows show the information flow among the various components described in this paper. Array shapes are shown in parentheses with s, number of sequences (Nseq in the main text); r, number of residues (Nres in the main text); c, number of channels (Jumper, Evans, Pritzel et al. 2021a).

Just to cite the improvement of the new architecture of AF2, it allows for jointly embedding of multiple sequence alignments (MSAs) and pairwise features. Moreover, AF2 has a new representation output and an associated loss that together allow for end-to-structure prediction.
The AF research teams does not submit that the system is capable of revealing underlying laws regulating protein folding. AF, however, seems to have reached important results concerning some kinds of proteins, especially those based on a strain of between 100 and 200 amino acids. Moreover, albeit the neural networks system distances the empirical link of evidence gathered from experimental data in the genomic database of proteins, it has the computational power to disclose the structure of the simulated object. In future, it may be capable of finding common patterns between the structures predicted. In any case, from a philosophical perspective, it is important to ask whether this kind of AI system can assist researchers in unveiling recurrent structures that could be defined as the laws governing protein folding. This discovery could improve even better the system solving the folding problem.

6. AlphaFold as a Simulation Model

In the last years, as the use of deep-learning neural networks has become pervasive in engineering and scientific areas,[9] scholars have focused correspondingly on the diffusion of simulation models as tools and outputs of neural network systems. What are simulation models[10] is then a crucial issue in the epistemology of models and the general philosophy of sciences.

A simulation model (SM) is a representation of a real or possible system, interacting with a determined environment, supported by computation techniques and expressed through visualization tools. It is a powerful instrument to represent, observe, study and manipulate to a higher degree of realism complex phenomena within a system. I submit that a model produced by AF is a kind of SM endowed with a degree of accuracy that was not available in the past, therefore improving the representational link between M and the related T. I submit that AF is a system architecture that produces SM of proteins' structures. We can divide AlphaFold into three main sectors: 1) AF as a complex neural-network system as a whole architecture; 2) AF sector sequences of algorithmic processing, the main blocks of the architecture; 3) AF's protein structure model as the output of the system.

As we know, the first stages of the system have to do with the analysis of the protein structure data contained in the database. In fact, in CASP the accuracy of the predicted structure is measured through the structure model obtained via experimental methods through X-ray crystallography and NMR spectroscopy. I claim that in each sector AF works as a kind of SM. According to CASP14 there are three relations to be noted:

1) The first between the real target system T and the experimental model, i.e. the relation between the real receptor-binding protein adhesin (Fig. 1) and the model resulting from the use of X-ray crystallography and NMR spectroscopy;

[9] See also Mitchell 2019; and Wooldridge 2021.
[10] See also Durán 2018, 2020; and Paronitti 2008, 2009.

2) The second between the experimental model and the simulation model AF, namely what is pictured in Fig1, the relation between the model obtained through experimental methods and the simulation model produced by AF;

3) The third between AF, the whole system architecture and the real target-system T, i.e. the real adhesin protein. The experimental and simulation success of these models is due to the relation they have with T. In the first case, the relation is obtained through experimental work which preserves the empirical link between observation and data manipulation. In the second case, the two different kinds of models are both successful representations of T, even though the simulation success entails a higher abstraction than the empirical link the experimental model holds in the first place. In the third case, AF as a whole architecture and the target system are not linked by an empirical relation, in so far as there is no direct observational contact as in the case of X-ray crystallography or NMR spectroscopy between the enquirer and the T. They are connected through the data manipulation and the simulation process binding the initial data, with the structure model in output.

Given the digital, computational, and algorithmic nature of the AF system, we can interpret it as an architecture producing simulation model (SM). There are mainly two types of simulation models: 1) SM is conceived as an implementation of models already existing; for example, aerospace engineers use SM of planes to test models they already have under specific circumstances like mechanical stress and weather conditions; 2) CS as models which have their own complexity and autonomy, the study of which is enhanced focusing on computer science and software engineering.[11] According to Durán (2021: 317), a simulation model (SM) is a "rich and complex structure that departs in important ways from standard models used in scientific research". Furthermore, Durán (2021) argues that the construction of the SM is possible because of a new methodology that is in place. He calls it *recasting*, and it consists of clustering a multiplicity of models into one fully computational SM. Think of it as the mashing-up of different models, also theoretical and mathematical, that could be implemented through deep-learning networks, with the specific aim to predict, in this case, the folding of proteins. To refine the terminology for the purposes of AF, we can call the methodology in place *reshaping*. AF begins with a set of data with empirical and experimental information, then through the intervention of programmers in adjusting the learning bias with respect to the desired output, using different integration modules, idealisations, and reshaping the data representation with the multiple sequence alignments MSA, according to cycles of implementation and integration, through the Evoformer and the Structure Module, we gain the visualization of the 3d geometry of the folding shape of the protein.

Not all the SM produced by AF are accurate representations of their T, especially the complex proteins are very hard to predict through the AF architecture as it is. Moreover, AF does not predict important aspects of protein structures as many ligands, metal ions and cofactors. Furthermore, the main limitation of AF is that the system predicts only a single state of the protein, and it is also hard to tell which state of the protein will be represented by the model (Perrakis and Sixma 2021). In fact, AF produces indeed SM with specific aims and empirical and theoretical assumptions and limitations, that must pass the abovementioned *accessibility condition* AC. Moreover, given the accuracy standard gained from the

[11] See also Symons and Alvarado 2019; Durán 2018; and Boyer-Kassem 2014.

experimental data, we can draw another requirement to be satisfied, the correctness condition (CC) for the proteins models:

> *Correctness condition*: SM represents correctly iff the accuracy qualification (AQ) is satisfied.

The AQ developed by CASP is a threshold for the correctness of the representation of SM. I take it as the level of approximation to reality the representation gains from the system through the work of modelers.

To conclude, AF consists not only of a complex and sophisticated computational implementation of the experimental models of proteins' structure determination, but it is a simulation model which is already changing the scenario of the computational and structure biology research areas.

7. Structure and Representation

I have advanced an interpretation of AF models as simulations. Thanks to the simulation power, modelers have greatly improved the representational capacity of models. Now I suggest a definition of the relation, refined through simulation, holding between the AF models and the objects they aim to represent:

> *Structural Dynamic Approximate Isomorphism*: a mapping that gathers through simulation even more information about the dynamic structure of T, so that the two systems (the model and T) approximately share the main structural features.

This definition pictures the ideal isomorphism between the model and the real protein which AF assumes as an implicit presupposition. It is a form of mapping since AF aims to visualize the shape of the protein as an image which can be navigated and observed in many aspects on a computer. The two systems should share the same features, represented one-to-one in the model: the individual folding units (domains), dynamic movements, contact matrix, ligands, and each polypeptide chain, and monomers, involved in multimers. Moreover, the two systems should share the same features under the same dynamics, i.e. the interactions of the domains in T should correspond in the mapping of the model. Given the limitation of AF, the definition assumes that the simulation model could be refined through time thanks to more and better information about the relevant features of the real proteins. The isomorphism between the two systems should regard the geometry as the information detected regarding the ligands and the folding units. In the case of protein folding the isomorphic relation is fundamental between the two systems, in so far as the protein shape is responsible for its function.

Why should the isomorphism be dynamic? One of the most important limitations of AF is that it predicts only a single state of a protein, but the aim of the AF researcher is to overcome this boundary. AF models are the peak of an important history of views about, and scientific representation of, proteins. In the last century structure biologists[12] shifted from the static view, according to which the protein models represented rigid structures, to the dynamic view:

[12] For a review of the history of structural biology, from the static to dynamic view, and a philosophical account of representation and explanation in the study of protein in structural biology, see Neal (2021).

> The study of how proteins serve the needs of a living organism is a curious case in which a method that yielded dramatic advances also led to a misconception. The method is X-ray crystallography […] The intrinsic beauty and the remarkable detail of the structures obtained from X-ray crystallography resulted in the view that proteins are rigid. This created the misconception, namely that the atoms in a protein are fixed in position (Karplus and McCammon 1986: 42).

The dynamic turn in protein representations owes a lot to thermodynamics. In fact, the dynamic analysis treats proteins as thermodynamic systems. The shift brought changes also to the structural concept. The old structural concept, coherent with the static view, is committed to the beliefs 1) that every protein has a rigid and static 3d structure and 2) that the protein structure alone determines protein function. The new dynamic concept of protein structure drops these commitments and adopts an inferential stance toward the proteins' structures, which are taken to be flexible, dynamic and constantly under structural fluctuations and mutations according to the environment and occurrent phenomena. Advocates of the dynamic concept are committed to the belief that dynamics and structures are relevant determinants of protein behavior and function (Neal 2021). The supporters of the dynamic concept suggest a wide range of experimental, theoretical and computational strategies to test the dynamic properties of proteins. AlphaFold researchers support the dynamic view of protein structure, well represented by accurate prediction models.

The motivation of AF is that biological research will be aided by the availability of an open-source determination structure database. The assumption underlying AF system and fostering this motivation is that simulation model structures entail an isomorphic relation with the target-protein. The protein may be in the real world, or a possible protein, or a protein mutation, whose structure is to be explored, in order to accomplish some specific functions, as in the case of PET depolymerization (Lu, Diaz, Czarnecki et al. 2022). AF model assumes that the dynamic view can be fostered through computational methods via deep-learning network architecture.

The AF system architecture is built to replicate the shape of the proteins according to their geometric features. The SM is apt to replace the representation of a protein given by the experimental procedures. The accuracy of the AF models is then grounded on the approximation to the structure of the real protein or to the functional structure of potential proteins. What best captures the conservation of information and geometric features between M and T is the notion of isomorphism. Related to protein structure prediction or drug discovery, AF researchers are therefore committed to a kind of isomorphism. On its basis, we can then define the representation relation:

> *Representation*: A scientific model M represents a T, which may be actual or potential, iff the dynamic structure of the model is approximately isomorphic to the structure of the T.

This kind of definition avoids some problems described in the structuralist conception of scientific representations.[13] According to Suárez (2003) and Downes (2009) isomorphism cannot ground the representation relation, because the former is characterized as reflexive and symmetrical, while the latter is not. Frigg and Nguyen (2017: 55) coin the requirement of directionality to account for this asymmetry. To

[13] See also Gelfert 2017; Frigg and Nguyen 2021.

answer these critics, let us recall that AF modelers do not aim at ideal models of proteins. The 100% GDT score is an ideal limit of research output, while the condition to be obtained is the standard of accuracy, i.e. AF models are accurate in so far as they represent their T, as an experimental representation of them would have done. The accuracy of AF models relies on the training the networks have got from the experimental data gathered. The isomorphic relation is approximate in the sense that the relation safeguards the correctness condition (CC).

Moreover, since the function of a protein depends on its folding, in the dynamics of interaction with the phenomena and molecules in the environment, there is a fundamental connection between the information it carries and the structure it takes once folded. Modeling such a dynamic structure allows us to understand the function of the protein. The isomorphism between the target-structure and the simulated or predicted structure is crucial to study, manipulate, and explore actual and possible functions of proteins. In so far as we need models to offer information about the target, the directionality of representation is then from model to target. It is indeed the asymmetry of the M-T relation that assures the accessibility condition (AC) that accurate models accommodate.

The isomorphic picture of the representational relation between the AF models and their T is one to take at face value if we want to develop a philosophical account of a breakthrough scientific advance such as AlphaFold.[14]

References

Alai, M. 2021a, "Scientific Realism and Further Underdetermination Challenges", *Axiomathes*, 31, 779-89.

Alai, M. 2021b, "The Historical Challenge to Realism and Essential Deployment", in Lyons, T.D. and Vickers, P. (eds.), *Contemporary Scientific Realism: The Challenge from the History of Science*, Oxford: Oxford University Press, 183-215.

Alai, M. 2023, "Scientific Realism, Metaphysical Antirealism and the No Miracle Arguments", *Found Sci*, 28, 377-400.

Bai, X.C., McMullan, G., and Scheres, S.H.W. 2015, "How Cryo-EM Is Revolutionizing Structural Biology", *Trends Biochem. Sci.*, 40, 49-57.

Boyer-Kassem, T. 2014, "Layers of Models in Computer Simulations", *International Studies in the Philosophy of Science*, 28, 4, 417-36.

Campbell, N.R. 1920, *Physics: The Elements*, 1st edition, Cambridge: Cambridge University Press.

Contessa, G. 2010, "Editorial Introduction to Special Issue", *Synthese*, 172, 2, 193-95.

Dill, K.A., Ozkan, S.B., Shell, M.S., and Weikl, T.R. 2008, "The Protein Folding Problem", *Annu Rev Biophys.*, 37, 289-316.

Downes, S. 2009, "Models, Pictures, and Unified Accounts of Representation: Lesson from Aesthetics for Philosophy of Science", *Perspective on Science*, 17, 4, 417-28.

Ducheyne, S. 2008, "Towards an Ontology of Scientific Models", *Metaphysica*, 9, 1, 119-27.

[14] I would like to thank Mario Alai and the two anonymous referees for their helpful comments along with the participants in the PRIN conference "The Manifest Image and The Scientific Image" held in Urbino, on the 20th and 21st of June 2022.

Durán, J.M. 2018, *Computer Simulations in Science and Engineering: Concepts—Practices—Perspectives*, Berlin: Springer.

Durán, J.M. 2020, "What Is a Simulation Model?", *Minds & Machines*, 30, 301-23.

Elowitz, M.B. and Lim, W.A. 2010, "Build Life to Understand It", *Nature*, 468, 7326, 889-90.

Fischer, E. 1894, "Einfluss der Konfiguration auf die Wirkung der Enzyme", *Berichte der deutschen Chemischen Gesellschaft*, 27, 3, 2985-93.

Fischer, E. 1906, *Untersuchungen über Aminosäuren, Polypeptide und Proteine*, Vol. 2, Berlin: Springer.

Frigg, R. and Nguyen, J. 2017, "Models and Representation", in Magnani, L. and Bertolotti, T. (eds.), *Springer Handbook of Model-Based Science*, Dordrecht: Springer, 49-102.

Frigg, R. and Nguyen, J. 2022, "Scientific Representation", *The Stanford Encyclopedia of Philosophy*, Zalta, E.N. (ed.), https://plato.stanford.edu/archives/win2021/entries/scientific—representation/

Gelfert, A. 2016*, How to Do Science with Models: A Philosophical Primer*, Berlin: Springer.

Giere, R. 1988, *Explaining Science*, Chicago: University of Chicago Press.

Giere, R. 1999, "Using Models to Represent Reality", in Magnani, L., Nersessian, N.J., and Thagard, P. (eds), *Model-Based Reasoning in Scientific Discovery*, New York: Plenum Publishers, 41-57.

Godfrey-Smith, P. 2006, "The Strategy of Model-Based Science", *Biology and Philosophy*, 21, 725-40.

Hesse, M.B. 1966, *Models and Analogies in Science*, Notre Dame: University of Notre Dame Press.

Hughes, R.I.G. 1997, "Models and Representation", *Philosophy of Science*, 64, 325-36.

Jaskolski, M., Dauter, Z., and Wlodawer, A. 2014, "A Brief History of Macromolecular Crystallography, Illustrated by a Family Tree and Its Nobel Fruits", *FEBS J.*, 281, 3985-4009.

Jumper, J., Evans, R., Pritzel, A. et al. 2021a, "Highly Accurate Protein Structure Prediction with AlphaFold", *Nature*, 596, 583-89.

Jumper, J., Evans, R., Pritzel, A. et al. 2021b, "Supplementary Information for Highly Accurate Protein Structure Prediction with AlphaFold", *Nature*, 596, 583-89.

Karplus, M. and McCammon J.A. 1986, "The Dynamics of Proteins", *Scientific American*, 254, 4, 42-51.

Kendrew, J.C. 1961, "The Three-Dimensional Structure of a Protein Molecule", *Scientific American*, 205, 6, 96-111.

Kendrew, J.C., Bodo, G., Dintzis, H. et al. 1958, "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis", *Nature*, 181, 662-66.

Knuuttila, T. 2021, "Epistemic Artifacts and the Modal Dimension of Modeling", *Euro Jnl Phil Sci*, 11, 65.

Le Bihan, S. 2016, "Enlightening Falsehoods: A Modal View of Scientific Understanding", in Grimm, S.R., Baumberger, C., and Ammon, S. (eds.), *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, London: Routledge, 111-35.

Lu, H., Diaz, D.J., Czarnecki, N.J. et al. 2022, "Machine Learning—Aided Engineering of Hydrolases for PET Depolymerization", *Nature*, 604, 662-67.

Mitchell, S. 2013, *Unsimple Truths: Science, Complexity, and Policy*, Chicago: University of Chicago Press.

Mitchell, M. 2019, *Artificial Intelligence: A Guide for Thinking Humans*, New York: Farrar, Straus and Giroux.

Moult, J. 2020, "AlphaFold: The Making of a Scientific Breakthrough", video interview, https://www.youtube.com/watch?v=gg7WjuFs8F4 (last access November 2023).

Moult, J. 2022, "CASP, Research Project", http://moult.ibbr.umd.edu/projects.htm (last access November 2023).

Morgan, M.S. and Morrison, M. 1999, "Models as Mediating Instruments", in Morgan, M.S. and Morrison, M. (eds), *Models as Mediators*, Cambridge: Cambridge University Press, 10-37.

Morrison, M. 2008, "Models as Representational Structures", in *Nancy Cartwright's Philosophy of Science*, Hartmann, S., Hoefer, C., and Bovens L. (eds.), Abingdon: Routledge, 67-88.

Neal, J.P. 2021, "Protein Structure, Dynamics and Function: A Philosophical Account of Representation and Explanation in Structural Biology", Doctoral Dissertation defended on July 27, 2021, https://d—scholarship.pitt.edu/41715/13/NealJP%20ETD%20History%20%26%20Philosophy%20of%20Proteins.pdf (last access November 2023).

Paronitti, G. 2008, *Epistemologia della simulazione: L'artificiale tra astrazione e realtà*, Morrisville: Lulu Press.

Paronitti, G. 2009, *Che cos'è la simulazione*, Roma: Carocci.

Pauling, L. and Corey R.B. 1951, "Atomic Coordinates and Structure Factors for two Helical Configurations of Polypeptide Chains", *Proc Natl Acad Sci USA*, 37, 235-40.

Pauling, L., Corey, R.B., and Branson, H.R. 1951, "The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain", *Proceedings of the National Academy of Sciences*, 37, 4, 205-11.

Perrakis, A. and Sixma, K.T. 2021, "AI Revolutions in Biology: The Joys and Perils of AlphaFold", *EMBO Reports*, 22, 11, 1-6.

Perutz, M.F. 1970, "Stereochemistry of Cooperative Effects in Haemoglobin: Haem-Haem Interaction and the Problem of Allostery", *Nature*, 228, 726-34.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C.T. 1960, "Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å: Resolution, Obtained by X-Ray Analysis", *Nature*, 185, 416-22.

Rosenblueth, A. and Wiener, N. 1945, "The Role of Models in Science", *Philosophy of Science*, 12, 4, 316-21.

Senior, A.W., Evans, R., Jumper, J. et al. 2020, "Improved Protein Structure Prediction Using Potentials from Deep Learning", *Nature*, 577, 706-10.

Suárez, M. 2003, "Scientific Representation: Against Similarity and Isomorphism", *International Studies in Philosophy of Science*, 17, 225-44.

Suárez, M. 2004, "An Inferential Conception of Scientific Representation", *Philos. Sci.*, 71, 5, 767-79.

Symons, J. and Alvarado, R. 2019, "Epistemic Entitlements and the Practice of Computer Simulation", *Minds and Machines*, 48, 4, 729.

Teller, P. 2001, "Twilight of the Perfect Model", *Erkenntnis*, 55, 393-415.

The AlphaFold Team 2020, "AlphaFold: A Solution to a 50-Year-Old Grand Challenge in Biology", https://www.deepmind.com/blog/alphafold—a—solution—

to—a—50—year—old—grand—challenge—in—biology (last access November 2023).

Thompson, M.C., Yeates, T.O., and Rodriguez, J.A. 2020, "Advances in Methods for Atomic Resolution Macromolecular Structure Determination", *F1000Res.*, 9, 667, 1-18.

Van Fraaseen, B. 2008, *Scientific Representation*, Oxford: Oxford University Press.

Weisberg, M. 2007, "Who Is a Modeler?", *The British Journal for the Philosophy of Science*, 58, 2, 207-33.

Wooldridge, M. 2021, *The Road to Conscious Machines: The Story of AI*, London: Pelican.

Wüthrich, K. 2001, "The Way to NMR Structures of Proteins", *Nat. Struct. Biol.*, 8, 923-25.