

**DESAFÍOS EPISTEMOLÓGICOS EN LA ERA DE LAS REDES
NEURONALES ARTIFICIALES: ABORDANDO SISTEMAS COMPLEJOS
DESDE UNA PERSPECTIVA COMPUTACIONAL**

***EPISTEMOLOGICAL CHALLENGES IN THE ERA OF ARTIFICIAL NEURAL
NETWORKS: ADDRESSING COMPLEX SYSTEMS FROM A
COMPUTATIONAL PERSPECTIVE***

ERICK RUBIO

erickrubio@mail.uniatlantico.edu.co

Universidad del Atlántico, Colombia - Grupo FiloComplex,
Sociedad Argentina de Análisis Filosófico, Argentina

LEANDRO GIRI

leandrogiri@gmail.com

Universidad Nacional de Tres de Febrero - Consejo Nacional de
Investigaciones Científicas y Técnicas (CONICET) - Grupo
FiloComplex, Sociedad Argentina de Análisis Filosófico,
Argentina

ANDRÉS ILCIC

ailcic@ffyh.unc.edu.ar

Instituto de Investigaciones Filosóficas de la Sociedad Argentina
de Análisis Filosófico, Consejo Nacional de Investigaciones
Científicas y Técnicas (IIF-SADAF-CONICET) - Grupo
FiloComplex, Sociedad Argentina de Análisis Filosófico,
Argentina

RECIBIDO: 22/08/2023

ACEPTADO: 24/11/2023

Resumen: En este artículo se explora un enfoque “manipulacionista” en la investigación científica, utilizando redes neuronales artificiales para abordar

sistemas complejos. Se destaca la capacidad de estas redes para resolver problemas y representar patrones de comportamiento en estos sistemas. Además, se enfatiza la importancia de superar las limitaciones de los métodos analíticos tradicionales mediante métodos computacionales y la visualización de grandes cantidades de datos. El artículo toma como ejemplo significativo el problema del plegamiento de proteínas y el éxito de AlphaFold2 en este campo. Se plantea la necesidad de explorar las implicaciones epistemológicas más amplias de las simulaciones computacionales en la investigación científica, considerando la relación entre precisión predictiva y comprensión de los sistemas complejos estudiados. La estructura del trabajo se divide en cuatro secciones, que incluyen el estudio de sistemas complejos, el problema del plegamiento de proteínas y una discusión filosófica sobre los modelos de simulación computacional y las redes neuronales artificiales; además de una sección de comentarios finales. En conjunto, se enfatiza la relevancia de este enfoque para abordar sistemas complejos y su impacto en la filosofía de la ciencia.

Palabras clave: sistemas complejos; redes neuronales artificiales; AlphaFold2; manipulacionismo; transparencia epistémica

Abstract: The article delves into a “manipulationist” approach in scientific research, utilizing artificial neural networks to tackle complex systems. It highlights the networks' capacity to resolve problems and portray behavioral patterns within these intricate systems. Moreover, it underscores the significance of surmounting the limitations posed by traditional analytical methods through computational techniques and the visualization of vast amounts of data. The article specifically focuses on the issue of protein folding and mentions the achievements of systems like AlphaFold2 in this domain. It highlights the necessity to explore the broader epistemological implications of computational simulations in scientific research, taking into account the interplay between predictive accuracy and the comprehension of the studied complex systems. The work is structured into four sections, encompassing the study of complex systems, the problem of protein folding, a philosophical discussion on computational simulation models and artificial neural networks, and concluding with final remarks. Overall, it underscores the significance of this approach in addressing complex systems and its profound impact on the philosophy of science.

Keywords: Complex systems; artificial neural networks; AlphaFold2; manipulationism; epistemic transparency

Introducción

En este artículo se explora la posibilidad de un abordaje metodológico científico de tipo «*manipulacionista*¹», en el que las técnicas de experimentación computacionales puedan ser una herramienta fructífera para el abordaje de *sistemas complejos*. En esta dirección, se examinará la capacidad de las simulaciones computacionales para resolver problemas concretos, a partir de la identificación de patrones de comportamiento de sistemas complejos y su representación científica.

¹ Es importante señalar que, si bien el término suele hacer referencia a cierta tradición en filosofía de la ciencia acerca de las explicaciones causales, con James Woodward a la cabeza, en el desarrollo de este trabajo no se hace mención, salvo en esta referencia, a James Woodward, tratando de tomar cierta distancia del problema de la explicación científica, y sobre todo de la causalidad. En el estudio de los sistemas complejos, dada sus características, no parece ser una estrategia favorable intentar un abordaje ligado a la causalidad. Piénsese, por ejemplo, en los impulsos nerviosos espontáneos que se dan en nuestro cerebro: el análisis de la causación interniveles genera enormes dificultades para dar cuenta de este fenómeno y otros por el estilo. El manipulacionismo de Woodward está pensado básicamente dentro de un contexto experimental empírico -de laboratorio-, donde se podría variar sistemáticamente las condiciones iniciales para obtener un mayor convencimiento de que ciertas condiciones son capaces de hacer aparecer un resultado particular, y ese proceso muestra que son esas y no otras las condiciones que lo causan. Woodward (p.e. 2003) está considerando el auge de la biología molecular a partir del desarrollo de técnicas e instrumentos para secuenciar y manipular por ejemplo moléculas *concretas* de ADN. A partir de una manipulación controlada de esta molécula se pueden obtener explicaciones causales, por ejemplo, de ciertas enfermedades. En cambio, como se verá, el manipulacionismo de nuestro enfoque actual sería principalmente de carácter *virtual*, algorítmico, abriendo incluso fuertemente la posibilidad a un automanipulacionismo que se distanciaría aún más de la propuesta woodwardiana. En el entrenamiento de una red neuronal podría modificarse el modelo subyacente con el que se estaría intentando abordar la dinámica del sistema.

Debido a sus características, los sistemas complejos difícilmente pueden ser tratados ni satisfactoria ni exclusivamente por métodos analíticos tradicionales, pues los mismos solo tienen un alcance limitado y, en el mejor de los casos, éxito muy parcial (Lenhard, 2016; Vicsek, 2002). El tipo de aproximación que ha logrado tratar con cierto éxito a los sistemas complejos está basado principalmente en el uso intensivo e imprescindible de métodos computacionales de diversa clase (Vicsek, 2002) y la representación visual de enormes cantidades de datos (Lenhard, 2017). Un análisis de lo anterior arroja que se debe renunciar, al menos parcialmente, a las pretensiones epistémicas normativas (y, quizás en algunos casos, incluso a las de comprensión) para poder tratar adecuadamente estos fenómenos (Lenhard, 2016). ¿Qué se entiende aquí por “éxito”? Como defendemos en este trabajo, no se trata de *transparencia epistémica* ni de entendimiento, mucho menos en términos del establecimiento de teorías generales; sino más bien de la solución de problemas particulares en contextos determinados, sobre los que se logra una comprensión parcial por la articulación de modelos distintos, cada uno de los cuales enfatiza algunas de las propiedades del sistema complejo bajo estudio.

En los últimos tiempos, las *redes neuronales artificiales* han sido objeto de discusión en la filosofía de la ciencia (Gao & Ganguli, 2015; Churchland & Sejnowski, 2016; Paninski & Cunningham, 2018; Chirimuuta, 2020), al ser consideradas como una forma de simulación computacional para abordar sistemas complejos; a veces como modelo propiamente y en otros casos como una técnica de procesamiento de datos. La dinámica de activación en las neuronas artificiales simularía (imitaría rudimentariamente), por ejemplo, la dinámica de las neuronas biológicas (Rumelhart et al., 1986). Suponiendo esto, lo que se buscaría básicamente sería extrapolar algunos resultados en el estudio dinámico de uno de los sistemas hacia el otro sistema en consideración, y la extrapolación se haría a

partir de los resultados obtenidos del empleo de ordenadores en el abordaje dinámico de uno de los sistemas (Hartmann, 1996). Este proceder, justamente en el terreno del que sea quizá el sistema complejo por excelencia, a saber: nuestro cerebro, está abriendo flancos promisorios de ataque para avanzar un poco más en su comprensión. Por ejemplo, Timothy P. Lillicrap, junto con Geoffrey Hinton, entre otros, en un interesante artículo titulado ‘Backpropagation and the brain’ (Lillicrap et al., 2020), exploran la posibilidad de que en ciertas áreas de la corteza cerebral se pueda dar una suerte de implementación de uno de los algoritmos más destacados en el entrenamiento de las redes neuronales artificiales como lo es el algoritmo de retropropagación del error.

Además, estas redes tienen la capacidad de “aprender” y adaptarse continuamente a partir de la experiencia, lo que las convierte en herramientas poderosas para la resolución de problemas de gran complicación, como el plegamiento de las proteínas, donde sistemas como *AlphaFold2* ha dado un gran avance recientemente.

En la literatura filosófica, se ha generado una controversia acerca de la justificación de las redes neuronales artificiales como herramientas para la investigación científica y la evaluación de sus resultados (Gao & Ganguli, 2015; Churchland & Sejnowski, 2016; Paninski & Cunningham, 2018). Algunos autores defienden su uso argumentando que proporcionan resultados precisos y útiles (Rumelhart et al., 1986; Lillicrap & Kording, 2019), mientras que otros cuestionan si son capaces de ofrecer una explicación adecuada de los fenómenos investigados (Gao & Ganguli, 2015; Pearl & Mackenzie, 2018).

En este contexto, Mazviita Chirimuuta (2020), por ejemplo, ha señalado que la investigación en decodificación —del cerebro en general y del neocórtex en particular— en la neurociencia

computacional revela un balance entre la precisión *predictiva*² y la capacidad de los modelos por simulación computacional —más concretamente por redes neuronales artificiales— para conferir comprensión. Según esta perspectiva, el progreso en la precisión predictiva no se traduce en un avance en la comprensión del sistema complejo objeto de estudio; en su caso, en una comprensión del cerebro. Ante la crítica de que el aprendizaje automático no produce un nivel aceptable de comprensión, algunos neurocientíficos (Bashivan et al., 2019, p. 1) sugieren redefinirla como la capacidad de predecir y controlar: en este caso, al cerebro. Chirimuuta defiende una noción no-factivista de la comprensión científica, que considera que los modelos reducen la complejidad natural a un tamaño manejable por los humanos mediante la abstracción e idealización.³ En este marco, la inteligibilidad de los modelos se ve comprometida cuando se despliegan algoritmos como las redes neuronales artificiales.

En otras palabras, la discusión acerca del valor de la comprensión como fin en sí mismo, no solo como medio hacia la predicción y el control, se presenta en el contexto del compromiso descrito por Chirimuuta.⁴ En definitiva, se plantea la necesidad de explorar las implicaciones de las simulaciones computacionales en la

² Sobre lo que podría considerarse el ‘giro predictivo’, particularmente en neurociencias, ver, por ejemplo: Weiskopf (2022)

³ Para un tratamiento comprehensivo del problema de la abstracción e idealización de los modelos científicos ver, por ejemplo: Frigg (2022, cap. 11).

⁴ La idea de comprensión suele estar asociada a la idea de encontrar, para los sistemas bajo estudio, los mecanismos causales (Durán, 2018). Sin embargo, para el caso de modelos que se construyen mediante enormes bases de datos y algoritmos de inteligencia artificial capaces de procesarlos en busca de correlaciones, se ha argumentado (Napoletani, Panza & Struppa, 2011; Mayer-Schomberg & Cukier, 2013) que la investigación correlacional debe reemplazar a la causal. En el presente trabajo, por motivos que iremos justificando durante su desarrollo, exploramos esta vía de razonamiento para los sistemas complejos y su análisis mediante simulaciones por redes neuronales.

investigación científica desde una perspectiva epistemológica más amplia; una perspectiva que tome fuertemente en consideración a los sistemas complejos.

El trabajo consta de cuatro secciones. En la primera se aborda el estudio de los sistemas complejos, que resulta inherentemente complicado por el surgimiento de propiedades emergentes, las cuales acarrearán dificultades epistemológicas y metodológicas para su abordaje predictivo y explicativo. También se trabaja sobre los fenómenos de autoorganización, en tanto hacen a otra característica típica de estos sistemas.

En la segunda se aborda el problema del plegamiento de proteínas y el avance significativo en su resolución logrado por AlphaFold, un sistema de redes neuronales artificiales desarrollado por la compañía DeepMind. El plegamiento de las proteínas es un desafío científico importante que ha interesado a investigadores en áreas como la biología, la bioquímica y la biomedicina durante décadas, y que nos servirá aquí de ejemplar paradigmático para señalar el modo en que la manipulación de parámetros relevantes de la dinámica del sistema a través de simulaciones “no clásicas” permite la identificación de patrones y correlaciones complejas difíciles de hallar por métodos analíticos y experimentales tradicionales.

En la tercera sección se lleva a cabo una discusión filosófica sobre los modelos de simulación computacional y las redes neuronales artificiales. Se argumenta que estos métodos de simulación, especialmente las redes neuronales, son herramientas poderosas para abordar sistemas complejos con una gran cantidad de elementos, pues permiten explorar las posibles consecuencias en las dinámicas de estos sistemas al alterar parámetros relevantes. Ello, argumentamos, conlleva consecuencias epistemológicamente relevantes que exploraremos con cierto detalle. Finalmente, presentaremos algunas conclusiones.

Aproximación a los sistemas complejos

Como suele ocurrir a la hora de dar una definición concreta del objeto que estudia un nuevo campo de actividad científica, se tiende a partir de definiciones poco concretas, con términos que tienen muchos referentes posibles, algunos incluso con la misma dificultad definicional que el campo que se quiere delimitar. Esto es lo que sucede al aproximarse al campo de los sistemas complejos. Una primera aproximación que responde a muchas de las inquietudes de quienes dieron forma al campo ha sido el estudio de aquellos sistemas que exhiben características emergentes no triviales y comportamientos autoorganizados (Mitchell, 2009, p. 13). Ahora bien, a la hora de caracterizar qué constituye una propiedad emergente o un comportamiento autoorganizado, se tiende rápidamente a brindar una suerte de definición negativa, es decir, una por la vía de “aquello que no”:

[Los] sistemas en los que los comportamientos organizados surgen sin un regulador o líder interno o externo son algunas veces llamados sistemas autoorganizados. Debido a que simples reglas producen comportamientos complejos *en cierta manera difíciles de predecir*, el comportamiento macroscópico de estos sistemas es algunas veces llamado emergente. (Mitchell, 2009, p. 13. nuestro énfasis)

Aunque aún no parece haber consenso sobre cómo definir a estos sistemas, sí se han logrado identificar ciertos rasgos distintivos, muchos de ellos en términos de la incapacidad de someterse a cierto tipo de descripción o de metodología de resolución de problemas característicos o paradigmáticos de los contextos de producción de conocimiento sobre otras clases de sistemas o fenómenos. Aquí nos

concentramos en dos aspectos íntimamente relacionados: la emergencia y la autoorganización.

Emergencia

El fenómeno de la complejidad suele ser visto como uno fuertemente asociado al de la emergencia (Bunge, 2004; Mitchell, 2009, 2004; Gell-Mann, 1998; Taylor, 2001); fenómeno en el que surgen propiedades globales irreductibles explicativamente a las propiedades de los (muchísimos) componentes (considerados bajo cierto nivel de organización) de un sistema, que dan lugar a ellas mediante sus múltiples interacciones *no lineales*.⁵

Las propiedades emergentes pueden tomarse entonces como propiedades de un objeto complejo que ni sus constituyentes ni sus precursores poseen (Bunge, 2004, p. 20; Johnson, 2003). Ello no quiere decir que sean propiedades que no se puedan abordar a partir de los componentes del sistema o su trayectoria. Más bien, no se pueden explicar a partir únicamente de las propiedades de sus partes constitutivas, puesto que se requiere conocer el tipo de interacción que están teniendo dichas partes, y, sobre todo, el conjunto de valores paramétricos –e *hiperparamétricos*⁶– que bajo estas

⁵ La no linealidad se refiere aquí a una relación entre variables donde el cambio en una de ellas no produce un cambio de magnitud proporcional en la otra. Es decir, no siguen una relación directamente proporcional. En cambio, la relación puede ser compleja y extremadamente difícil de predecir.

⁶ La noción de *hiperparámetro* juega un rol muy importante en el tipo de enfoque aquí propuesto; a saber, uno basado principalmente en redes neuronales artificiales. En este sentido debemos diferenciar la estructura hiperparamétrica del modelo (cantidad de capas, cantidad y tipo de neuronas, etc.), de la estructura de parámetros que surge (o emerge) luego del entrenamiento del modelo. Por citar algunos ejemplos: para identificación o clasificación de imágenes, se modelan redes neuronales convolucionales; para predicciones en series de tiempo, redes recurrentes; para procesamiento de textos, redes recurrentes o Transformers; todas ellas con una estructura hiperparamétrica particular muy diferente y que facilita

interacciones no lineales presentan algunas variables consideradas relevantes para la dinámica del sistema complejo en estudio. Es ahí donde radica una dificultad esencial en el análisis de los sistemas complejos: en disponer de este conjunto de interacciones y de los valores paramétricos que bajo tales interacciones presentan ciertas variables implicadas en la descripción de la dinámica del sistema, ya que estos suelen ser sistemas dinámicos no deterministas, que, bajo ciertas condiciones, algunas veces favorecidas por el azar, presentan nuevas asociaciones o combinaciones de sus elementos, lo que conlleva a que el conjunto de interacciones varíe. Esta variación en sus interacciones viene acompañada regularmente de nuevos comportamientos que suelen dar lugar a nuevas propiedades del sistema (motivo por el cual muchos de ellos se dicen adaptativos (Holland, 1992). Ello podría llevarnos a considerar que el quid de las propiedades emergentes está en las interacciones no lineales y en los valores de ciertos parámetros –e hiperparámetros, según el enfoque basado en redes neuronales artificiales aquí propuesto–, que bajo tales interacciones presentan ciertas variables ligadas a los elementos constituyentes de estos sistemas complejos; propiedades que no son distributivas, sino globales (Bunge, 2004, p. 29).

Las propiedades globales antedichas pueden ser, por ejemplo, de tipo *funcional* –como la que se desprende de la estructura tridimensional de las proteínas–, *estructural* –como el procesamiento de información de tipo jerárquico en el neocórtex– u *organizacional* –como la que presenta una bandada de estorninos (*Sturnus vulgaris*) en busca de alimento o en defensa de otras aves

enormemente el proceso de aprendizaje del sistema (ver Feurer & Hutter, 2019; Yang & Shami, 2020). El enfoque que desarrollamos más adelante en este trabajo sugiere que los hiperparámetros pueden verse como “parámetros de parámetros” con los cuales se puede ganar conocimiento del sistema gracias a crear una representación de alto nivel que permite manipular al menos parcialmente la dinámica del sistema a fin de encontrar propiedades robustas para su descripción.

rapaces (Rescher, 1998, p. 9). El conocimiento, aun detallado y preciso, de las propiedades de los componentes de un sistema vistos de manera aislada nos resulta insuficiente para poder dar cuenta de estas propiedades en sentido global. Esto sucede pese a considerar que es a partir de dichos componentes (considerados bajo cierto nivel de organización o abstracción), combinados o asociados, que dichas propiedades tienen lugar, y que es a partir de la interacción de estos que se pueden definir y detectar casos específicos de dichas propiedades globales.

Las mencionadas propiedades del sistema bien sean de tipo funcional, estructural u organizacional, requieren de un gran conocimiento adicional al que se tenga de las propiedades de los componentes para intentar dar cuenta de ellas; además del desarrollo de técnicas y estrategias no analíticas, sino más bien aproximativas, implementadas iterativamente y retroalimentadas mediante muchísimo ensayo y error. La emergencia, característica notoria de la complejidad, siendo principalmente un fenómeno ontológico (en el sentido de Humphreys, 1997, 2016), representa serias dificultades epistemológicas, particularmente en sentido metodológico, para abordarla predictiva y explicativamente. Sin embargo, a propósito de la diferenciación entre estructura hiperparamétrica y paramétrica de un sistema de aprendizaje automático, podemos afirmar que, si bien es cierto que las propiedades globales emergentes no pueden explicarse ni preverse a priori partiendo de la estructura paramétrica que será aprendida por el sistema (dado que, en la práctica, es imposible prever cuáles serán los parámetros que surgirán de la iteración), creemos que sí es posible prever ciertas características de las propiedades globales a partir de la estructura hiperparamétrica del modelo.

De acuerdo con cómo estructuramos los hiperparámetros del sistema (la estructura marco previa a la iteración de aprendizaje), podemos tener cierta noción de las propiedades globales emergentes

que esperamos observar. Por ejemplo, si utilizamos redes neuronales artificiales de tipo convolucional, sabremos que las propiedades globales (o regularidades aprendidas por el sistema) se centrarán en la detección de formas o disposición/relación espacial de/entre los datos de entrada (Feurer & Hutter, 2019). En cambio, si diseñamos una red de tipo recurrente (al estilo Hopfield, una LSTM o incluso un Transformer), las regularidades aprendidas se centrarán en el orden u ocurrencia temporal de los datos. Incluso podemos diseñar este tipo de redes de manera tal que la cantidad de regularidades o propiedades aprendidas podamos conocerla de antemano (por ejemplo, utilizando una última capa densa con una cantidad de neuronas igual a la cantidad de propiedades que estimamos predecir).

Podemos también pensar que si lo que deseamos obtener son propiedades de tipo estructural jerárquico, una estructura hiperparamétrica de redes convolucionales con *pooling* podría ayudar en dicho propósito (dado que van “abstrayendo” en cada capa *pool* las regularidades más relevantes o generales, ver Feuerer & Hutter, 2019).

Por tanto, si bien no podemos predecir o explicar con total transparencia el porqué de las propiedades globales emergentes dada la estructura del sistema, sí podemos prever ciertas características esperadas de las mismas. Volveremos a este importante punto más adelante.

Autoorganización

Cuando se habla de sistemas autoorganizados se está pensando en sistemas que adquieren cierto orden sin el control o la dirección central de mando. El concepto de orden no es fácil de caracterizar; sin embargo, una manera de acercarnos a él puede ser mediante el concepto de información.

‘Información’, en los sistemas complejos, y en general en los sistemas físicos, es opuesto a ‘entropía’ (Hidalgo, 2015, p. 24). En ese sentido, diremos que los sistemas complejos, centrándonos en este momento en el orden que exhiben, son tanto peculiares como poco comunes (Hidalgo, 2015, p. 23), ya que sus muchísimos elementos constituyentes presentan disposiciones físicas con un altísimo grado de información, cuyas rutas de acceso para llegar a ellas, ateniéndose a las constricciones que operan sobre tales elementos, son extremadamente limitadas. Estos estados ricos informacionalmente se dan en extremadamente muy pocos microestados, por lo que la entropía de estos estados es bastante baja.

Pensemos, por ejemplo, en una proteína. Sabemos que es una configuración espacial específica de su secuencia de aminoácidos la que la hace funcional. El espectro de configuraciones o plegamientos posibles, dadas las constricciones fisicoquímicas que operan sobre los aminoácidos constitutivos de una proteína, es extremadamente amplio (Levinthal, 1968). Sin embargo, de todas esas posibles configuraciones, o de todas esas posibles maneras en las que podrían plegarse los aminoácidos encadenados, la funcionalidad de la proteína está reducida a una configuración particular. Decimos entonces que es esa configuración particular la que contiene un alto contenido de información. Es rica en información con respecto al fondo de las conformaciones no funcionales.

Aquí, el concepto de complejidad efectiva resulta de gran utilidad. Murray Gell-Mann lo define como “la longitud de una descripción concisa de las regularidades [de un sistema]” (Gell-Mann, 1998, p. 66). Es decir, se está pensando en una descripción comprimida de las relaciones conspicuas que los elementos constituyentes del sistema presentan en esos estados ricos informacionalmente. Se está pensando en la descripción más concisa del conjunto de patrones identificables entre las interacciones que estos elementos presentan.

Podemos decir entonces que los sistemas que exhiben comportamientos autoorganizados son sistemas que presentan una disposición de sus elementos constitutivos con alta complejidad efectiva y baja entropía; son sistemas cuyos elementos constituyentes adquieren cierta disposición física de manera espontánea con un alto contenido de información. El comportamiento autoorganizado de sistemas complejos es uno en cuya disposición de los elementos que dan lugar a él podemos encontrar ciertos patrones, pero también cierta dosis de aleatoriedad. Curiosamente, estos estados se suelen dar en lo que se conoce como el borde del caos (Nicolis & Prigogine, 1989). Y curiosamente también, ahí encuentran una suerte de equilibrio entre la flexibilidad y la rigidez de sus comportamientos que les resultan indispensables para adaptarse (Kauffman, 1993).

Complejidad y redes neuronales artificiales: el problema del plegamiento de proteínas y el gran adelanto en su solución por parte de AlphaFold

El problema del plegamiento de las proteínas ha interesado sobremanera por décadas a los científicos en el área de la biología, la bioquímica, la biomedicina, etc. Tanto así, que desde el año 1994 ha venido realizándose un concurso (*Critical Assessment of Techniques for Protein Structure Prediction -CASP-*), donde se premia al grupo o centro de investigación que logre *predecir* la estructura o configuración precisa de un amplio número de proteínas a partir de su secuencia de aminoácidos. Hasta el momento, el concurso se ha llevado a cabo en quince ocasiones. En dos de ellas, (CASP13: 2018 y CASP14: 2020), ha participado DeepMind, que es un laboratorio de inteligencia artificial de enorme prestigio mundial, fundado en 2010 y adquirido por Google en 2014. En ambas

ocasiones resultó ganador,⁷ y con un buen margen de ventaja con respecto a otros equipos que emplearon otras aproximaciones numéricas al plegado de proteínas. De hecho, el margen obtenido en su segunda participación -con AlphaFold2- fue abrumadoramente amplio. Según Venki Ramakrishnan, Premio Nobel de química y presidente de la Royal Society

Este trabajo computacional representa un avance sorprendente en el problema del plegamiento de proteínas, un gran desafío de 50 años en biología. Han transcurrido décadas antes de lo que las personas en el campo habían estimado. Será emocionante ver las muchas formas en que cambiará fundamentalmente la investigación biológica (Ramakrishnan, citado en Jumper et al., 2020)

El problema del plegamiento de las proteínas es importante porque las proteínas son moléculas esenciales para la vida y juegan un papel crítico en muchos procesos biológicos, incluyendo la regulación de la actividad celular, la comunicación celular, la defensa contra enfermedades y la producción de energía (Evans et al., 2018). La forma tridimensional en la que se pliegan las proteínas determina su función biológica, y las proteínas mal plegadas pueden ser la causa de una amplia variedad de enfermedades, incluyendo neurodegenerativas como el Alzheimer, el Parkinson, o el mal de Huntington (Suárez, 2005, p. 243). Además, la resolución del problema del plegamiento de las proteínas también es importante desde una perspectiva científica más amplia. Es uno de los problemas más importantes y difíciles de la biofísica y la biología molecular, y su solución puede tener implicaciones importantes en

⁷ En el sitio web de la conferencia (predictioncenter.org) se pueden comparar los modelos presentados por los distintos equipos a lo largo de la historia de la competición. CASP15 se llevó a cabo durante 2022 y los resultados mejoraron considerablemente con respecto a los años anteriores, en su mayoría tras la aplicación de métodos de aprendizaje automático similares a las arquitecturas introducidas por AlphaFold 1 y 2.

muchos campos, incluyendo la medicina, la ingeniería de proteínas y la biotecnología.

La red neural artificial de aprendizaje profundo utilizada por el equipo de DeepMind para obtener un modelo tridimensional de la proteína, a partir de la secuencia de aminoácidos que la constituye, se conoce como AlphaFold. La misma es heredera de las capacidades ingenieriles de la empresa, que ya había ostentado un éxito abrumador con las versiones AlphaGo y AlphaZero, desarrolladas para jugar al Go y al ajedrez respectivamente: juegos en los que se evidencia que una exploración de todas las posibles partidas para determinar estrategias conducentes a una victoria resulta imposible. El abanico de posibilidades de combinaciones de movimientos es prácticamente infinito (Evans et al., 2018). Al igual que en AlphaFold, claramente, la búsqueda no puede ser exhaustiva. La misma se realiza mediante un árbol de búsqueda de Markov;⁸ esto es, un árbol de búsqueda aleatorio. Si bien se realiza con cierta dosis de aleatoriedad, esta es guiada con la intervención del algoritmo del descenso del gradiente⁹ (*gradient descent*), junto al de *backpropagation* (Nielsen, 2015) o retropropagación del error. Este

⁸ En un árbol de búsqueda de Markov, se representa un problema de toma de decisiones secuenciales, donde cada nodo del árbol representa un estado del sistema y cada arista representa una acción que se puede tomar en ese estado. Además, cada nodo contiene información sobre la recompensa asociada con ese estado y la probabilidad de transición a otros estados. El árbol de búsqueda de Markov se construye de manera recursiva, explorando el espacio de posibles acciones y estados futuros a partir del estado actual. Una de sus ventajas es su capacidad para modelar y resolver problemas complejos de toma de decisiones secuenciales en entornos inciertos. (Ver, por ejemplo, Puterman, 1990)

⁹ Más concretamente, el algoritmo del descenso del gradiente estocástico (SGD) y el descenso del gradiente adaptativo (Adam). Este último, en lugar de usar una tasa de aprendizaje constante como ocurre en el descenso del gradiente estándar, adapta la tasa de aprendizaje en función de los momentos del gradiente de los parámetros, lo que de alguna manera facilita o permite que esta se adapte automáticamente a la topología del modelo y a la naturaleza de los datos.

algoritmo permite ir minimizando de manera iterativa la función de error, o el coste del modelo, seleccionando un subconjunto aleatorio de los datos para su entrenamiento (*mini-batch*) y analizando en cada caso la distancia con una solución conocida. Lo que se hace con este algoritmo es básicamente moverse en el espacio de búsqueda (traducido a la superficie de la función de coste) intentando encontrar mínimos locales mediante la manipulación de los valores de los parámetros considerados en dicho modelo. Así, la esencia del algoritmo consiste en actualizar los pesos del modelo en la dirección del gradiente negativo utilizando en cada actualización o iteración un subconjunto aleatorio de los datos seleccionados para su entrenamiento.

Para el caso de AlphaFold, los parámetros considerados para la modelización son las distancias que separan a los aminoácidos que constituyen a la proteína y los ángulos de torsión de los enlaces químicos de dichos constituyentes (aminoácidos). Afirman los creadores de AlphaFold que

[...] Las propiedades que predicen nuestras redes son: (a) las distancias entre pares de aminoácidos y (b) los ángulos entre los enlaces químicos que conectan esos aminoácidos. [...]

Entrenamos una red neuronal por separado para predecir una distribución de distancias entre cada par de residuos en una proteína. Estas probabilidades se combinaron en una puntuación que estima cuán precisa es una estructura de proteína propuesta. También entrenamos una red neuronal separada que usa todas las distancias en conjunto para estimar qué tan cerca está la estructura propuesta de la respuesta correcta.

Usando estas funciones de puntuación, estuvimos en capacidad de buscar en el paisaje de proteínas para encontrar estructuras que coincidieran con nuestras predicciones. Nuestro primer método se basó en técnicas comúnmente utilizadas en biología estructural, y se reemplazaron repetidamente piezas de una estructura proteica con nuevos fragmentos. Entrenamos una red neuronal generativa para

inventar nuevos fragmentos, que se utilizaron para mejorar continuamente la puntuación de la estructura proteica propuesta.

El segundo método optimizó los puntajes a través del descenso del gradiente –una técnica matemática comúnmente utilizada en el aprendizaje automático para realizar pequeñas y crecientes mejoras–, lo que resultó en estructuras altamente precisas. (Evans et al., 2018, s/p)

Se intenta mostrar con lo anterior que un punto clave en el gran adelanto de resolución de este problema, que lleva algo más de cincuenta años de planteado, consistió en la manipulación y exploración de las consecuencias de dicha manipulación, vía modelo y simulación computacional, de los parámetros estimados relevantes en la dinámica del sistema. Esta exploración, dada la enorme cantidad de combinaciones de valores involucrados en el fenómeno emergente a explicar –la funcionalidad de la proteína–, no puede darse de manera exhaustiva, sino más bien aleatoriamente. A partir del conocimiento de las propiedades fisicoquímicas de los aminoácidos constituyentes de una proteína particular resulta impensable deducir la estructura o configuración específica que la hace funcional, requiriéndose entonces vías alternativas.

Las redes neuronales artificiales son una herramienta que permiten dar cuenta de la interacción no lineal de muchísimos elementos. Esto representa una estrategia con la que se podría dar un tipo de aproximación no analítica de fenómenos emergentes como los que interesan en el estudio de los sistemas complejos. Se habla de aproximaciones no analíticas en el sentido de que en ellas no se parte estrictamente de un análisis de los contenidos de conocimiento de los elementos involucrados en el fenómeno –emergente– a explicar para poder dar cuenta de este. Se parte más bien de una suerte de manipulación o intervención, por medio de simulaciones computacionales de dichos elementos involucrados, hasta obtener una *correlación* entre el fenómeno a explicar y los valores de ciertos parámetros considerados relevantes en la dinámica del sistema y que

se pueden codificar bajo los parámetros de otro sistema sobre el que la manipulación es más factible.

En este sentido, el enfoque de simulación computacional permite identificar patrones y correlaciones que podrían pasar desapercibidos con un enfoque analítico tradicional, o bien directamente ser imposibles de lograr. El carácter “manipulacionista” de este tipo de aproximaciones radica en la posibilidad de manipular a gran escala los parámetros de la simulación para obtener resultados específicos y correlacionarlos con los datos del fenómeno a explicar (Vicsek, 2002), de hecho, guiando la manipulación hacia una mejor representación de la relación entre entradas y salidas. Así, se podría decir que la importancia de las correlaciones en la epistemología radica en la posibilidad de identificar patrones y relaciones emergentes que podrían ser difíciles de detectar mediante métodos analíticos tradicionales, y que la simulación computacional es una herramienta útil para identificar estas correlaciones y generar nuevas hipótesis.

Es importante resaltar que la limitación de un acercamiento expresamente analítico al fenómeno de la emergencia, y en general al fenómeno de la complejidad, va más allá de la dificultad de encontrar esquemas generales aplicables a estos fenómenos. En *principio*, mediante el sistema de ecuaciones diferenciales parciales para representar la función de onda de Schrödinger se podría determinar la configuración molecular que pretendamos. Sin embargo, con estas ecuaciones diferenciales parciales, a partir de más de diez partículas, no es ya sólo que una solución analítica resulta prácticamente imposible, sino que, además, una solución por aproximación numérica resulta en extremo difícil, aun con el poder de cálculo que presentan hoy en día los computadores más poderosos de los que disponemos (Laughlin & Pines, 2000). Piénsese que sólo estamos considerando algo más de diez partículas. La situación para el problema del plegamiento de proteínas es extremadamente

inmanejable desde la función de onda de Schrödinger, ya que en una proteína intervienen cientos de miles de partículas. Sin embargo, como se ha podido observar con el ejemplo de AlphaFold, a partir de la identificación de ciertos parámetros y de la exploración, posibilitada por simulaciones computacionales de las consecuencias de la manipulación de un amplio conjunto de valores paramétricos, ha resultado una estrategia conducente a resultados positivos.

AlphaFold, u otros sistemas exitosos cuyos resultados sorprendentes no eran previsibles a partir de las propiedades individuales de sus componentes (neuronas), posee una estructura de hiperparámetros diseñada muy cuidadosamente para tal fin. En este diseño podemos ver que sí hubo una intervención ‘analítica’ más o menos precisa: por ejemplo, una decisión clara de diseño está implicada en la selección específica como parámetros de las distancias que separan a los aminoácidos que constituyen a la proteína. Además, incluso hasta el algoritmo del descenso de gradiente tiene una justificación totalmente analítica: ¿por qué utilizar las derivadas parciales para actualizar los parámetros en cada iteración y no utilizar otro método, o simplemente el azar puro? Claro está que en el diseño hiperparamétrico de cualquier sistema de aprendizaje automático existe una importante intervención de conocimiento de tipo teórico y técnico, tanto del dominio de aplicación (en nuestro ejemplo, la bioquímica) como de la ingeniería computacional.

También es claro que las propiedades globales específicas que finalmente resulten del aprendizaje del sistema no son explicables ni predecibles con total transparencia de modo analítico, al menos en la mayoría de los casos. Esto es lo que suele denominarse el “problema de la caja negra” (Durán & Jongsma, 2021; Zednik, 2019). Sin embargo, sí existe la posibilidad de prever sus características mediante el diseño de hiperparámetros de un modo analítico. Además, si bien las propiedades globales no pueden preverse de

antemano, más allá de sus características generales, en algunos casos sí pueden explicarse a posteriori: los casos donde los resultados de la red neuronal pueden entenderse bajo ciertas explicaciones observables en el sistema del mundo real que el sistema artificial está modelando (en dichos casos, podríamos rastrear la “lógica” seguida por la red, aunque hacerlo quizás no nos diga mucho en última instancia del fenómeno que estamos intentando comprender, dado su comportamiento emergente).

Discusión filosófica sobre modelos de simulación computacional y redes neuronales artificiales

En este artículo se pretende sostener que, en el abordaje de sistemas complejos, constituidos de cantidades abrumadoras de elementos, algunos métodos de simulación computacional (aquí nos enfocamos en las redes neuronales) han resultado una poderosa herramienta, principalmente porque han permitido una exploración –no sin cierta opacidad– manipulacionista de posibles consecuencias en sus dinámicas a partir de ciertas alteraciones en alguno o algunos de los parámetros estimados relevantes para intentar dar cuenta de dicha dinámica.

El asunto es que la exploración facilitada por las redes neuronales no puede ser exhaustiva, como bien se intentó evidenciar con el ejemplo del problema del plegamiento de proteínas. Incluso para los potentes ordenadores de los que se disponen hoy día, explorar exhaustivamente las posibles consecuencias en la dinámica de los sistemas complejos de alterar los valores de ciertos parámetros estimados relevantes en el intento de dar cuenta de dicha dinámica bajo cierto modelo resultaría una tarea de nunca acabar. Si bien se establecen determinados criterios para la exploración, aun dentro del dominio de valores paramétricos resultante la exploración

exhaustiva resultaría algo improcedente. Así, una vez establecidos ciertos criterios para la exploración en cuestión que trazan ciertas “rutas de tránsito” preliminarmente plausibles —que bien podríamos tratar de heurísticas—, el movimiento exploratorio se da más bien de manera aleatoria.

En definitiva, en las redes neuronales artificiales de aprendizaje profundo —que son las de nuestro interés—, se presenta una situación epistemológica novedosa, y es una suerte de opacidad frente a lo que está ocurriendo en detalle en la búsqueda exploratoria de los valores de ciertos parámetros que se *correlacionan* con el fenómeno complejo/emergente a abordar.¹⁰

Con las redes neuronales artificiales se sabe qué es lo que está ocurriendo en términos generales en el despliegue del modelo subyacente que está siendo ejecutado —mediante la estructura

¹⁰ Lo que se tiene en el ámbito de las redes neuronales es una opacidad epistémica más profunda que la teorizada por Paul Humphreys (2004; 2009) quien la definía del siguiente modo: “un proceso es esencialmente epistémicamente opaco a [un agente cognitivo] X si y sólo si es imposible, dada la naturaleza de X, para X conocer todos los elementos epistémicamente relevantes del proceso (Humphreys, 2009, p. 618). Para Humphreys, las simulaciones computacionales son epistémicamente opacas por la incapacidad humana de conocer la evolución de todos los elementos epistémicamente relevantes de los procesos computacionales mientras ocurren. Si bien la opacidad epistémica no sería exclusiva de los procesos computacionales, la naturaleza de la tecnología involucrada (Hardware y Software) la haría especialmente acuciante allí, toda vez que muchas de las operaciones que allí suceden lo hacen a un nivel muy bajo, lejos del usuario. Humphreys menciona la opacidad con relación al uso de simulaciones computacionales utilizadas para estudiar *targets* del mundo real. Sin embargo, las simulaciones que analiza Humphreys son auditables (su código es accesible, y fue programado manualmente por expertos humanos), cosa que no sucede con las simulaciones con redes neuronales, donde las ‘lógicas’ utilizadas para llegar a los modelos finales, como veremos, son mucho menos transparentes toda vez que los parámetros se logran en procesos que involucran aprendizaje automático y el procesamiento de enormes bases de datos. Sin embargo, aun con esta opacidad, veremos que es posible obtener conocimiento útil también de las redes neuronales.

hiperparamétrica—, pero es prácticamente imposible determinar de antemano cuáles serán los valores finales de los parámetros aprendidos, dado que es muy difícil determinar a priori a qué mínimo local convergerá el sistema en su proceso de aprendizaje por el descenso del gradiente y, mucho más, cuáles serán los muchísimos parámetros aprendidos que conducirán a este mínimo error local. Es más, se diseña una programación de una naturaleza muy distinta a aquella en la que se dan las instrucciones precisas de qué hacer, más bien aquí se dota de suficiente variabilidad para que el algoritmo encuentre una serie de valores (como los numerosos parámetros que representan la intensidad de la conexión entre las neuronas de la arquitectura) que optimizan una métrica con respecto a un conjunto de datos. En esta búsqueda, el azar está implicado en la inicialización de variables y en otros procesos tales como la conformación de lotes aleatorios en el descenso del gradiente estocástico. Pero, más allá del azar implicado, lo que resulta prácticamente imposible es la determinación o previsión *a priori* de los mínimos locales existentes y a cuál de dichos mínimos la red convergerá: es la complejidad del problema y no el aspecto parcialmente no determinista lo que impide un tratamiento analítico tradicional. El azar, de hecho, es un recurso necesario, similar a como lo es en los métodos llamados Monte Carlo (Galison, 1996).

Los modelos de redes neuronales artificiales son una herramienta útil para resolver problemas computacionales complejos, pero su opacidad epistémica es un tema de discusión, por ejemplo, entre los expertos en la neurociencia, y, sobre todo, recientemente, en el terreno de la filosofía de la ciencia (Barak, 2007; Chirimuuta, 2020; Durán & Jongsma, 2021; Lipton, 2006). Aunque los constructores del modelo conocen muchas características de su arquitectura y funcionamiento interno, la forma exacta en que una red llega a sus predicciones o clasificaciones a menudo es bastante opaca para ellos. Esta opacidad se refleja en el término “caja negra”, que se utiliza

para describir dispositivos o piezas de código que transforman entradas en salidas sin proporcionar ninguna indicación del método detrás de esta operación (Chirimuuta, 2020, p. 3). En cuanto a esto, Omri Barak, por ejemplo, señala que “el aprendizaje automático nos proporciona niveles mayores de rendimiento, acompañados de un aumento paralelo en la opacidad” (2007, p. 5). Sin embargo, no sería del todo preciso afirmar que las redes neuronales artificiales son *completamente* opacas, ya que los creadores del modelo tienen conocimiento de varias características de su estructura y funcionamiento interno. Además, conocemos la conectividad completa de la red, la dinámica de cada neurona individual, la regla de plasticidad utilizada para entrenar la red y, de hecho, toda la experiencia de desarrollo de la red (Chirimuuta, 2020, p. 3).

El problema no menor que surge consiste en poder ofrecer una caracterización de los grados relativos de transparencia y opacidad exhibidos por diferentes modelos. Frente a esto, un punto de partida interesante puede ser la noción de ‘*inteligibilidad*’. De acuerdo con De Regt y Dieks (2005, p. 143), la falta de inteligibilidad impide que un predictor perfecto de tipo caja negra, conocido como un oráculo, sea considerado una teoría científica válida. Según su perspectiva, una teoría científica debe ser comprensible, lo cual implica la capacidad de entender cómo se generan las predicciones y desarrollar un entendimiento de las consecuencias que la teoría tiene en situaciones específicas. La inteligibilidad no se limita a un componente psicológico adicional, sino que es fundamental para la habilidad de los científicos para utilizar las teorías (Chirimuuta, 2020, p. 3).

De Regt reelabora una idea presentada originalmente por Werner Heisenberg y respaldada luego por Richard Feynman para caracterizar la inteligibilidad:

Criterio de Inteligibilidad (CIT): una teoría científica T (en una o más de sus representaciones) se considera comprensible para los científicos

(en el contexto C) si pueden reconocer las consecuencias cualitativas características de T sin necesidad de realizar cálculos exactos (De Regt, 2014: p. 33; también De Regt y Dieks, 2005, p. 151 y ss; Chirimuuta, 2020, pp. 3, 4).

En nuestro caso de análisis, frente a esta suerte de opacidad epistémica ligada fuertemente al despliegue de modelos de redes neuronales artificiales, podemos oponer tres elementos que pueden ayudar a echar luz a esta caja negra. Por un lado, las estructuras hiperparamétricas, a las que ya nos hemos referido, si bien no nos permiten prever los mínimos locales y las propiedades emergentes exactas subyacentes, sí nos habilitan a prever ciertas características que poseerán dichas propiedades aprendidas por la red (como su cantidad, su capacidad de detectar regularidades espaciales, temporales o de otro tipo, etc.). Por otro lado, si bien no es posible determinar de antemano qué regularidades detectará la red luego de su aprendizaje (si así fuera, no sería necesaria su utilización para el problema estudiado), en algunos casos sí es posible entender *a posteriori* dichas regularidades. Claro que no siempre es posible entender las “decisiones” adoptadas por la red y, muchas veces, estas son completamente opacas a nuestro entendimiento analítico, tanto *a priori* como *a posteriori*.

Respecto a lo que explayamos, podemos observar que en el caso en que las regularidades o propiedades globales aprendidas por la red puedan ser entendidas *a posteriori*, la opacidad epistémica podría considerarse solo aparente o al menos relativa al sistema computacional empleado, no con respecto al fenómeno sobre el que se ha aprendido algo gracias al método. De hecho, en el caso de AlphaFold, se puede argumentar que si bien no nos proporciona entendimiento del plegado mismo de las proteínas, su éxito predictivo con respecto a las estructuras puede constreñir el espacio de modelos posibles para el mecanismo de plegamiento,

proveyéndonos de esa manera nuevo conocimiento, quizás indirecto, de cómo estudiar luego el fenómeno subyacente.

Pensemos, además, que dicha opacidad sería similar a la que tenemos respecto a los razonamientos que realiza nuestro propio cerebro: podemos explicar a posteriori la lógica de nuestros pensamientos, pero nada sabemos de los procesos físicos detallados que siguieron nuestros miles de conexiones neuronales para llegar a ellos. Por lo tanto, el desconocimiento exacto de la estructura paramétrica (y de cómo arribó a ella la red en su entrenamiento) no entraña una opacidad epistémica *per se*; o, al menos, en todo caso, entraña la misma opacidad epistémica que nuestro desconocimiento de los procesos neuronales exactos (habitualmente denominados como “sub-personales”) que realiza nuestro cerebro cuando pensamos aun del modo más analíticamente riguroso. Podríamos agregar que la red artificial tiene la ventaja epistémica, sobre nuestro cerebro, de que podemos diseñar y conocer su estructura hiperparamétrica con total detalle y prever ciertas características de las propiedades emergentes resultantes (aun antes de que resulten, es decir, *a priori*), precisamente gracias a la manipulabilidad.

Por último, creemos que existe un tercer elemento que clarifica la opacidad epistémica atribuida a los sistemas computacionales: las imágenes que se van generando en sus iteraciones juegan un papel muy importante. Llegados a este punto conviene diferenciar dos tipos de situaciones que se dan con las imágenes generadas a partir de simulaciones computacionales, particularmente con aquellas generadas por redes neuronales artificiales. En primer lugar, es importante señalar que la cantidad de imágenes que se suelen generar por métodos de simulación computacional en general es sumamente grande, por lo que intentar identificar por parte nuestra algún tipo de patrón a partir de ellas es también una tarea improcedente. Sería una tarea de nunca acabar. En el caso, por ejemplo, de las redes neuronales artificiales, la identificación de patrones a partir de esa

enorme cantidad de imágenes ocurre regularmente por parte de ellas mismas (Evans et al., 2018).

En una de las etapas en el entrenamiento de AlphaFold, la red genera, basándose en ciertas técnicas propias de la bioquímica,¹¹ una enorme cantidad de imágenes que representan mapas de distancia entre los aminoácidos constituyentes de la proteína de la cual se pretende predecir su estructura tridimensional específica. Sobre estos mapas de distancia la red se entrena en la identificación de ciertos patrones. Con este entrenamiento, entre otras cosas, la red autoajusta ciertos parámetros para generar nuevos mapas de distancia más discriminados. Todo este proceso resulta opaco para la supervisión humana detallada. Todo ello implica que dentro de las simulaciones computacionales mediante redes neuronales artificiales se da el caso de la generación de imágenes y procesamiento de estas sin supervisión humana, algo que no ocurre en otros contextos de investigación científica.

¹¹ En primer lugar, los investigadores de DeepMind se valen del extenso *dataset* del que se dispone, desde hace ya un par de décadas, sobre la secuencia genómica de varias especies. Es decir, se valen básicamente del extenso número de proteínas de las que se conoce su secuencia de aminoácidos, a través de una técnica que se conoce como ‘Multiple Sequence Alignment’ (MSA) (Senior et al. 2019), con la que se busca principalmente alinear secuencias de aminoácidos constituyentes de proteínas emparentadas o ligadas evolutivamente. Es decir, se busca alinear secuencias homólogas.

La idea con este alineamiento es identificar patrones entre dichas secuencias o, más concretamente, correlaciones, las cuales de alguna manera ayuden a estimar la distancia a la que se encuentran dos aminoácidos entre sí. Por ejemplo, si al alinear varias secuencias de aminoácidos, constituyentes correspondientemente de varias proteínas ligadas evolutivamente de distintas especies (secuencias homólogas) se logra identificar bajo algún patrón de alineamiento una fuerte correlación entre dos aminoácidos, esto representaría un fuerte indicio de que esos dos aminoácidos se hallan a una distancia cercana el uno del otro. Opera aquí un principio de parsimonia, entendiendo que la selección natural busca conservar estructuras funcionales.

La otra situación que cabe resaltar en cuanto a las imágenes generadas por simulación computacional es aquella en la que éstas figuran como puentes de acceso o de observación para los investigadores a lo que ocurre “semiautónomamente” en la ejecución de estos modelos que involucran entrenamiento y búsqueda exploratoria con cierta dosis de aleatoriedad. Es decir, las imágenes en esta situación terminan siendo nuestro vínculo expedito con el modelo computacional en ejecución. Esto constituye la primera vez en nuestra historia que contemplamos una especie de autoridad epistémica artificial luego, por ejemplo, del éxito predictivo que vienen sosteniendo las redes neuronales artificiales (ver p. ej. Chirimuuta, 2020). Debido a que las imágenes que se generan a partir de sus ejecuciones terminan siendo en varios casos nuestro punto de acceso a ellas, podría considerarse que esta situación en términos de visualización resulta novedosa, no tanto por la novedad de la visualización científica en sí, que tiene larga data, sino por la escala en la que estas técnicas cuya arquitectura está inspirada en la corteza visual del cerebro puede crear y manipular representaciones visuales. Por este motivo, posibilita la contemplación de múltiples relaciones entre elementos al mismo tiempo (una visión holística del sistema). Muchas simulaciones, desde el inicio de la computación, proveen imágenes que mejoran la comprensión de los sistemas analizados. Lo interesante aquí es que estas imágenes puedan dar cuenta de sistemas complejos, y también que sean las propias redes neuronales las que analicen a las imágenes que generaron para extraer la mayor cantidad de información útil de las mismas. Así, estas nuevas imágenes generadas por simulaciones permiten a los investigadores arrojar un poco de luz a la opacidad epistémica inherente al análisis de sistemas complejos y del uso de redes neuronales en general, aunque la transparencia total resulte un ideal inalcanzable.

Por supuesto, pretensiones epistémicas como la transparencia podrían considerarse un poco desplazadas cuando de abordar la complejidad se trata. Frente al estudio de sistemas complejos, una comprensión o entendimiento analítico es reemplazado en gran medida por una comprensión o entendimiento manipulacionista enfocado quizás más bien en lograr resolver un problema heurísticamente y ofrecer predicciones. Ello ocurre con claridad en el ejemplo de AlphaFold y el problema del plegamiento de las proteínas, especialmente en lo que respecta al subproblema de la predicción de estructuras. Las simulaciones computacionales por redes neuronales permiten hacerles frente a las barreras de la complejidad antes que eludirlos, pues presentan una alternativa frente a la frustración de nuestra incapacidad de una comprensión analítica de toda realidad compleja (Lenhard, 2016, p. 20).

Conclusión

La discusión sobre las redes neuronales artificiales y su capacidad para proporcionar una explicación adecuada de los fenómenos que están siendo investigados y su capacidad para explicar adecuadamente fenómenos complejos está estrechamente relacionada con la noción de hiperparámetros y la opacidad epistémica que rodea a estos modelos. Si bien los hiperparámetros pueden proporcionar cierta claridad al permitir prever ciertas características de las propiedades aprendidas por la red, la incertidumbre persiste debido a la imprevisibilidad de las regularidades que la red detectará después del entrenamiento. Aunque en ocasiones es posible entender estas regularidades a posteriori, en muchos casos, las decisiones de la red siguen siendo opacas tanto antes como después del aprendizaje.

En el contexto del problema del plegamiento de proteínas, se destaca cómo las redes neuronales artificiales, como AlphaFold2 de DeepMind, son una valiosa herramienta para la investigación científica en la era de la complejidad. Este tipo de problema, que ha sido objeto de estudio durante décadas, presenta desafíos significativos debido a su complejidad y a la gran cantidad de variables involucradas. La comprensión analítica tradicional ha sido extremadamente difícil de lograr en este caso.

En lugar de buscar una comprensión detallada y analítica del problema, AlphaFold2 utiliza una comprensión manipulacionista, basada en el procesamiento masivo de datos por parte de la red neuronal para producir predicciones precisas y útiles. Aunque esto implica renunciar a la virtud epistémica clásica de la transparencia, se considera un precio justo a pagar para obtener conocimientos sobre sistemas complejos que de otro modo serían difíciles de abordar.

En resumen, la discusión sobre las redes neuronales artificiales en la investigación científica refleja una tensión entre la necesidad de una comprensión profunda y detallada de los fenómenos y la utilidad práctica de las herramientas que nos permiten manipularlos y predecirlos eficientemente. Los hiperparámetros desempeñan un papel relevante en intentar entender estas redes, pero aun así, la opacidad epistémica persiste, lo que destaca la importancia de considerar enfoques manipulacionistas en la resolución de problemas complejos con estas herramientas tecnológicas. Podemos estimar que parte de la dinámica científica por venir se centrará en la articulación de distintas clases de modelos sobre un mismo fenómeno

Referencias

- Barak, O. (2017). Recurrent neural networks as versatile tools of neuroscience research. *Current Opinion in Neurobiology*, 46, 1–6.
- Bunge, M. (2004). *Emergencia y Convergencia. Novedad Cualitativa y Unidad del Conocimiento* (Rafael González del Solar, Trad.). Gedisa. (Obra original publicada en 2003)
- Chirimuuta, M. (2020). Prediction versus understanding in computational enhanced neuroscience. *Synthese*, 199(4).
- Churchland, P. S., Sejnowski, T. J. (2016). Blending computational and experimental neuroscience. *Nature Reviews Neuroscience*, 17, 667–668.
- Durán, J. M. (2018). *Computer Simulations in Science and Engineering: Concepts -Practice -Perspectives*. Springer Nature Switzerland.
- Durán, J. M., Jongsma, K. R. (2021). Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *Journal of Medical Ethics*, 47(5), 329-335.
- Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., ..., Senior, A. (2018). *De novo structure prediction with deep-learning based scoring*.
- Feurer, M., Hutter, F. (2019). Hyperparameter Optimization, en: Hutter F., Kotthoff, L., & Vanschoren, J. (eds.) *Automated Machine Learning: Methods, Systems, Challenges* (pp. 3-33). Cham: Springer Nature Switzerland.
- Frigg, R. (2022). *Models and Theories: a Philosophical Inquiry*. Routledge.
- Galison, P. (1996). Computer Simulations and the Trading Zone, en: Galison P., & Stump, J. D. (eds). *The Disunity of Science: Boundaries, Context and Power* (pp. 118-157). Stanford University Press.

- Gao, P., Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Current Opinion in Neurobiology*, 32, 148–155.
- Gell-Mann, M. (1998). *El Quark y el Jaguar: aventuras en lo simple y lo complejo* (A. García y R. Pastor, Trad.). Tusquets. (Obra original publicada en 1994).
- Goldenfeld, N. & Kadanoff, L. P. (1999). Simple lessons from complexity. *Science*, 284: 87- 89.
- Hartmann, S. (1996). The World as a Process: Simulations in the Natural and Social Sciences, En: Hegselmann, R., Mueller, U., Troitzsch, K.G. (Eds). *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View* (pp. 77-100). Kluwer.
- Hawkins, J., Blakeslee, S. (2005). *On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*. St. Martin's Griffin.
- Hidalgo, C. (2015). *Why Information Grows. The Evolution of Order, from Atoms to Economies*. New York: Basic Book.
- Holland, J. H. (1992). Complex Adaptive Systems. *Daedalus*, 121(1), 17–30.
- Humphreys, P. (1997). Emergence, Not Supervenience. *Philosophy of Science*, 64, 337- 345.
- Humphreys, P. (2004). *Extending Ourselves. Computational Sciences, Empiricism, and Scientific Method*. Oxford University Press.
- Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, 169, 615-626.
- Humphreys, P. (2016). *Emergence: A Philosophical Account*. Oxford University Press.
- Johnson, S. (2003). *Sistemas emergentes. O qué tienen en común hormigas, neuronas, ciudades y software*. Fondo de Cultura Económica.

- Jumper, J., Evans, R., Pritzel, A. Green, T., Figurnov, M., Ronneberger, O., ..., Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596: 583-589.
- Kauffman, S. A. (1993). *The Origins of Order: Self-organization and Selection in Evolution*. Oxford University Press.
- Laughlin, R. B., Pines, D. (2000). The Theory of Everything. *Proceedings of the National Academy of Sciences*, 97(1), 28-31.
- Lenhard, J. (2016). Computer Simulation, en: Humphreys, P. (ed.) *The Oxford Handbook of Philosophy of Science* (pp. 803-824). Oxford University Press.
- Lenhard, J. (2017). License to Explore: How Images Work in Simulation Modeling, en: Ammon S., & Capdevila-Werning, R. (eds). *The Active Image. Architecture and Engineering in the Age of Modeling* (pp. 233-254). Springer.
- Levinthal, C. (1968). Are there pathways for protein folding? *Journal de Chimie Physique*, 65, 44-45.
- Lillicrap, T., Kording, K. (2019). *What does it mean to understand a neural network?*
- Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J., Hinton, G. (2020). Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6), 335-346.
- Mitchell, M. (2009). *Complexity. A Guided Tour*. Oxford University Press.
- Napoletani, D., Panza, M., Struppa, D. C. (2011). Agnostic science. Towards a philosophy of data analysis. *Foundations of Science*, 16(1), 1-20.
- Nicolis, G., Prigogine, I. (1989). *Exploring Complexity. An Introduction*. W. H. Freeman & Company.
- Nielsen, M. (2015). *Neural Networks and Deep Learning*.
- Pearl, J., Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

- Puterman, M. (1990). Markov Decision Processes. En *Handbooks in Operations Research and Management Science vol 2* (pp. 331-434). Elsevier.
- Rescher, N. (1998). *Complexity. A Philosophical Overview*. Transaction Publishers.
- Rumelhart, D., Hinton, G., Williams, R. (1986). Learning representations by back-propagation errors. *Nature*, 323, 533-536
- Senior A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., ..., Hassabis, D. (2020). Improved protein structure prediction using potentials from Deep learning. *Nature*, 577: 706-710.
- Taylor, M. C. (2001). *The Moment of Complexity. Emerging Network Culture*. Chicago: The University of Chicago Press.
- Vicsek, T. (2002). The Bigger Picture. *Nature*, 418: 131.
- Weiskopf, D. (2022). The Predictive Turn in Neuroscience. *Philosophy of Science* (5).
- Woodward, J. (2003). *Making Things Happen. A theory of causal explanation*. New York: Oxford University Press.
- Yang, L., Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415: 295-316
- Zednik, C. (2019). Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence. *arXiv:1903.04361 [cs]*.