ASRI
ARTE Y SOCIEDAD
REVISTA DE INVESTIGACIÓN
EN ARTES Y HUMANIDADES DIGITALES

VISIÓN ARTIFICIAL Y PERCEPCIÓN HUMANA: CORRESPONDENCIAS Y DIVERGENCIAS

ARTIFICIAL VISION AND HUMAN PERCEPTION: CORRESPONDENCES AND DIVERGENCES

Andrés Pachón Arrones
Universidad de Coimbra

DOI: 10.33732/ASRI.6891

Recibido: (02 12 2025)
Aceptado: (02 12 2025)

Cómo citar este artículo
Pachón Arrones, A. (2025). Visión artificial y percepción humana: correspondencias y divergencias. *ASRI. Arte y Sociedad. Revista de investigación en Arte y Humanidades Digitales*, (29), e6891. Recuperado a partir de https://doi.org/10.33732/ASRI.6891

Resumen
Esta investigación parte de la correlación que se estableció entre la neurociencia y la computación con la llegada del conexionismo a las investigaciones en Inteligencia Artificial (IA), una relación que daría lugar al actual *deep learning* y, en particular, a las *convolutional neural networks* (CNN) para visión artificial. La hipótesis que plantea este artículo es que esta relación trajo consigo una *neuromitología* que sustenta el discurso mediático de la IA contemporánea, por

Abstract
This research stems from the correlation established between neuroscience and computation with the emergence of connectionism in Artificial Intelligence (AI) research—a relationship that would give rise to contemporary deep learning and, in particular, to convolutional neural networks (CNNs) for computer vision. The hypothesis proposed in this article is that this relationship brought with it a neuromythology underlying the contemporary AI media narrative, according to which we are close to achieving a

el cual estamos cerca de alcanzar una *superinteligencia* computacional de carácter humano. Tomando como caso de estudio las CNN, se analizarán las relaciones y divergencias entre visión artificial y percepción humana, con el objetivo de desmitificar este discurso.

Palabras clave
IA, *deep learning*, CNN, neuromitología, abducción.

human-like computational superintelligence. Using CNNs as a case study, the article analyzes the relationships and divergences between computer vision and human perception, with the aim of demystifying this narrative.

Keywords
AI, deep learning, CNN, neuromythology, abduction.

Introducción

La mejora de las técnicas del *deep learning* en los últimos años y su masiva implementación han alimentado un discurso mediático por el cual se presupone un avance ilimitado para la Inteligencia Artificial (IA). Gurús tecnológicos como Elon Musk, Sam Altman o Mark Zuckerberg acreditan que a través del *deep learning* evolucionaremos hacia una IA de carácter humano, una *superinteligencia* (Altman, 2024) que será capad de lograr nuevos avances científicos por sí sola (Kaput, 2024), dando respuesta a algunos de los mayores desafíos de la humanidad, como el cambio climático o la cura del cáncer (Seisdedos, 2023). Este discurso se sustenta en una mitología cuyos antecedentes se remontan a la máquina de Turing y a su articulación con las ciencias cognitivas *computacionalistas* (Pachón, 2025). Sin embargo, a partir de los años ochenta se produce un punto de inflexión decisivo para la mitología contemporánea de la IA: se trata de la llegada del modelo conexionista, resultado de un fructífero intercambio entre la neurociencia y la computación. Este modelo presentaba una nueva forma de representar la realidad en un sistema artificial, de una manera más rica que con la IA simbólica, una vez que ya no era necesario traducir los objetos y eventos del mundo en símbolos deficientes de significado que requiriesen una definición adicional (Aleksander, 2002, pp. 250-251).

El conexionismo encontró su punto álgido con el desarrollo de las redes neuronales multicapa, base del actual *deep learning*, siendo uno de sus grandes logros las redes para reconocimiento, clasificación y generación de imágenes. Las redes más implementadas en la actualidad para tareas de visión computacional son las *convolutional neural networks* (CNN). Estas redes son empleadas en reconocimiento facial, conducción autónoma o el diagnóstico médico a través del análisis de imágenes (rayos X, resonancias magnéticas, etc.). Además, estos modelos serán la base de los modelos generativos de imágenes, como son las *generative adversarial networks* (GAN). Aunque el propósito del *deep learning* no es reproducir la estructura anatómica del cerebro, encontramos una excepción en las CNN, una vez que estas redes se inspiraron en el funcionamiento del córtex visual, lo que a su vez permitió que estas redes se usaran también como modelo computacional para respaldar determinadas investigaciones sobre los procesos neurobiológicos implicados en la visión. En este sentido, la neurocientífica teórica Grace Lindsay destaca la influencia mutua de la computación y la neurobiología para el estudio de la percepción visual:

> *Neuroscientists and computer scientists forged a long history of collaborating in their attempts to understand the fundamental questions of vision. The study of vision —of how patterns can be found in*

> *points of light— is full of direct influence from the biological to the artificial and viceversa. The harmony may not have been constant: when computer science embarked on methods that were useful but didn't resemble the brain, the fields diverged. And when neuroscientists dig into the nitty-gritty detail of the cells, chemicals and proteins that carry out biological vision, computer scientists largely turn away. But the impacts of the mutual influence are still undeniable, and plainly visible in the most modern models and technologies* (Lindsay, 2022, p. 154).

Esta particularidad convierte a las CNN en un excelente caso de estudio, ya que su historia técnica, basada en la relación entre computación y neurobiología, desafía nuestra resistencia al discurso que predice una *superinteligencia* de carácter humano basada en el *deep learning*.

Metodología

Propongo, en primer lugar (sección 1), analizar las correspondencias entre neurociencia y computación que surgieron durante el desarrollo de los modelos computacionales de visión artificial, comenzando por el *Neocognitrón* que dio origen a las CNN. En segundo lugar (sección 2), se examinarán las diferencias ontológicas entre la visión artificial y la percepción visual humana. Para ello, se abordarán las limitaciones técnicas del *deep learning* como simulador de la actividad neurobiológica, y se realizará un análisis semiótico que contraponga la inferencia abductiva propia de la percepción visual al método inductivo que emplean las CNN para clasificar imágenes.

Desarrollo de la investigación

1. Correspondencias entre neurociencia y computación en la percepción visual

Algo que parece tan sencillo como encontrar una silla en una sala, se presenta como un funcionamiento biológico complicado de comprender, ya que la luz de la sala, los objetos que están a su alrededor, la forma de la silla, su posición, su color, etc. pueden variar, ofreciendo millones de combinaciones que afectan a la forma en que los fotones de luz llegan a la retina y, aun así, todos ellos significan para nosotros que hay una silla. En palabras de Lindsay (2022) *the visual system somehow finds a way to solve this many-to-one mapping in less than a tenth of a second* (p. 153).

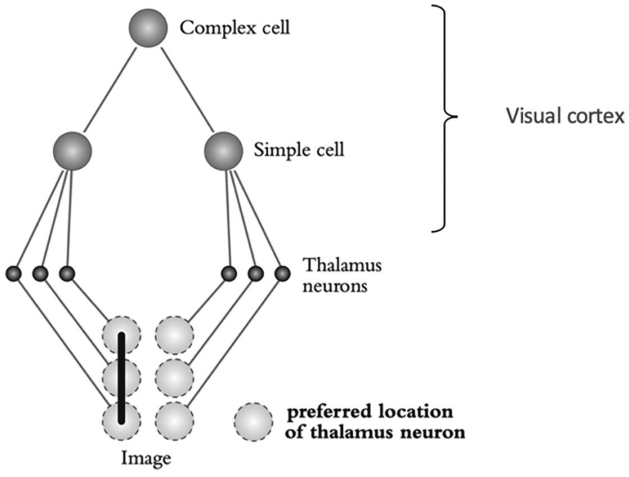
Un momento fundamental para la comprensión de este funcionamiento biológico fue el descubrimiento realizado por los neurofisiólogos Hubel & Wiesel en 1962, quienes ubicaron por primera vez dos tipos de células nerviosas principales (a las que denominaron *células simples* y *células complejas*) en la corteza visual primaria de los gatos, lo cual se demostró extrapolable al resto de mamíferos, entre ellos los humanos. Hasta este descubrimiento, tan solo conocíamos el funcionamiento y las conexiones neuronales entre el ojo y la corteza visual del cerebro. A saber: tenemos una lámina de células fotorreceptoras en la retina, denominadas conos y bastones, las cuales son sensibles a la luz visible reflectada por los objetos. Cada una de estas células indica la presencia o ausencia de luz en cada localización de la escena en cada momento, mandando una señal en forma de actividad eléctrica. Las células fotorreceptoras estimuladas envían esta señal a un otro conjunto de neuronas ubicadas en el tálamo, una estructura que se encuentra situada en el centro del encéfalo y que funciona como una estación intermedia por donde

pasa la información sensitiva antes de llegar al córtex cerebral. Como indica Lindsay (2022), las neuronas del tálamo parecen responder mejor a puntos simples — *either a small area of light surrounded by dark or a small area of dark surrounded by light* — (p. 165), siendo que cada neurona tiene una ubicación preferida donde debe estar el punto para que esta sea estimulada. Las neuronas del tálamo que sean activadas mandarán una señal a las células nerviosas de la corteza visual primaria —también denominada corteza V1—, una zona ubicada en el lóbulo occipital de la corteza cerebral primaria que está en la región posterior del cerebro.

Lo que Hubel & Wiesel descubrieron fue que estos impulsos enviados por las neuronas del tálamo llegan a un primer conjunto de células de la corteza visual que sus descubridores denominaron *células simples*, las cuales se activan con líneas oscuras o claras, siendo que cada una de estas células tiene una orientación preferida además de una ubicación preferida. De acuerdo con Lindsay (2022) *a neuron won't respond to just any line that shows up in its favourite location. Horizontal-preferring neurons require a horizontal line, vertical-preferring neurons require vertical lines, 30-degree-slant-preferring neurons require 30-degree slanted lines, and so on and so on* (p. 167). De esta forma, cada *célula simple* se activará con mayor intensidad cuando el conjunto de puntos enviados por las células del tálamo conforme una línea que se acerque a su orientación preferida (Figura 1). Lindsay resume este proceso de la siguiente manera:

> *Inputs to a neuron in the primary visual cortex therefore must come from a set of thalamus neurons wherein each one represents a dot in a row of dots. (...) neurons in the primary visual cortex listen for the activity of neurons in the thalamus that make up their preferred line* (Lindsay, 2022, p. 168).

Figura 1: Representación de las etapas de la percepción visual.



Fuente: Lindsay, 2022.

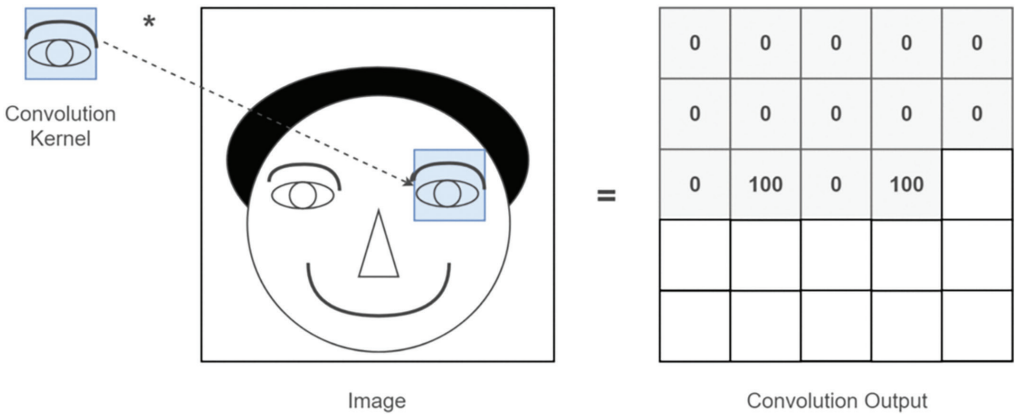
El segundo tipo de células que ubicaron Hubel & Wiesel en el córtex visual fueron las denominadas *células complejas*. Cada una de estas neuronas será estimulada por un conjunto de células simples activadas que tengan la misma orientación preferida, aunque sus ubicaciones sean diferentes. De esta forma, las *células complejas* maximizan la orientación de los *inputs* de las *células simples* independientemente de la exactitud de su ubicación, agrupando así la información recibida, lo cual será fundamental para el funcionamiento del sistema visual (ver Figura 1).

Lindsay ejemplifica la importancia de las *células complejas* de la siguiente manera:

> *If we want to know if the letter 'A' is in front of us, a little bit of jitter in the exact location of its lines shouldn't really matter. Complex cells are built to discard jitter. The discovery of complex cells provided an additional piece of the puzzle as to how points of light become perception. In addition to the feature detection done by simple cells, pooling of inputs across space was added to the list of computations performed by the visual system* (Lindsay, 2022, p. 168).

La descripción ofrecida por Hubel & Wiesel sobre el funcionamiento neuronal del sistema visual fue la base para el desarrollo de las CNN, siendo la primera de estas redes convolucionales el *Neocognitrón* desarrollado en los 80 por Kunihiko Fukushima, quien se basó en los procesos descritos por Hubel & Wiesel. Este modelo de red neuronal realizaba la clasificación de números escritos a mano, con la particularidad de que podía identificar los números sin importar su ubicación en la imagen de entrada, una vez que un mismo número escrito a mano por diferentes personas dará como resultado variaciones en la forma de ese mismo número, lo cual permitía demostrar el acercamiento del *Neocognitrón* a las capacidades de la visión biológica. En primer lugar, para introducir la imagen del número escrito a mano en el *Neocognitrón*, cada punto que conformaba la imagen del número era traducida a una señal de entrada, unos valores binarios que simulaban la entrada de la luz en el tálamo. Para calcular los valores de las entradas en la siguiente capa de la red —que corresponden a los valores de las células simples—, Fukushima usó una cuadrícula numérica, técnicamente denominada *filtro*, que se superponía a las diferentes ubicaciones de la imagen. Cada uno de estos *filtros* representaba una línea de orientación específica. De esta forma, el valor de cada *célula simple* se calculaba multiplicando los valores del *filtro* al sumatorio de los valores de la imagen de entrada en cada ubicación (Figuras 2, 3 y 4). Al deslizar un *filtro* por todas las ubicaciones de la imagen se genera un conjunto de *células simples* que tienen la misma orientación preferida, pero con diferentes ubicaciones (capa Us1 de la Figura 3). Esta población de *células simples* se encuentra en la capa convolucional de la red. Para activar los valores de la siguiente capa —que corresponden a las *células complejas*—, un otro algoritmo agrupa aquellas *células simples* que, en ubicaciones próximas de la imagen, tienen una misma orientación. Esto permitía detectar patrones más complejos de la imagen (capa Uc1 de la Figura 3). Esta capa de *células complejas* se denomina técnicamente *capa de submuestreo* (Figura 4).

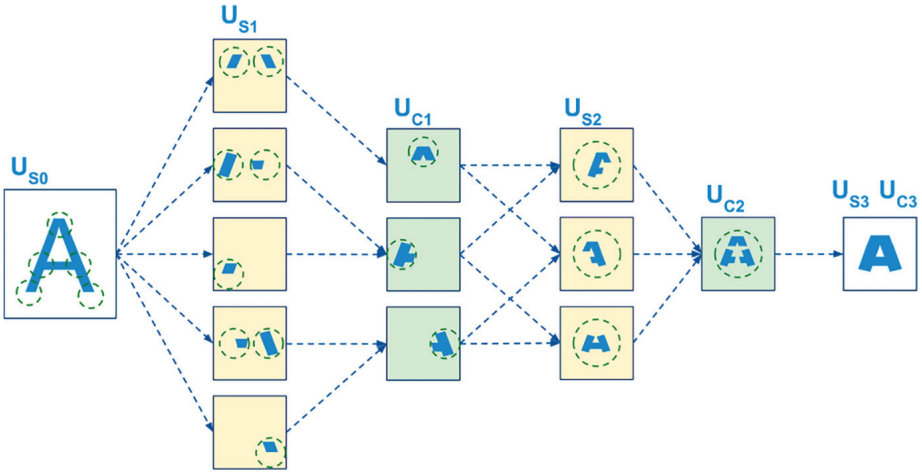
Figura 2: Representación del funcionamiento de un *filtro*. Al desplazarse por las distintas partes de la imagen, el *filtro* (que aquí se corresponde con el dibujo de un ojo) extrae una determinada característica (un valor) que, en el caso de las *células simples* del *Neocognitrón*, se refiere a la orientación de una línea (ver Figura 3).



Fuente: Cuartas, 2021.

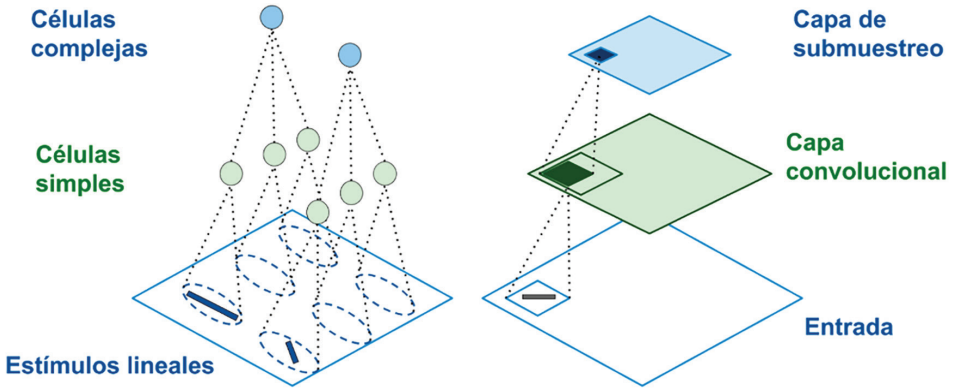
Universidad Rey Juan Carlos
ASRI - Arte y Sociedad. Revista de Investigación
Núm. 29. Año 2025. ISSN: 2174-7563
CUVPAC Grupo de investigación CULTURA VISUAL & PRÁCTICAS ARTÍSTICAS CONTEMPORÁNEAS

Figura 3: Representación del funcionamiento del *Neocognitrón*, donde se puede ver el resultado de los filtros convolucionales en las células simples (Us1) y del submuestreo en las células complejas (Uc1).



Fuente: Negrón, 2020.

Figura 4: Relación entre el sistema visual biológico establecido por Hubel & Wiesel y las capas de una CNN.



Fuente: Negrón, 2020.

Ante la falta de una descripción neurobiológica del resto del sistema visual que permitiera explicar la detección de patrones más complejos, Fukushima decidió repetir esta estructura, aplicando una segunda ronda de *neuronas simples* que recibían su *input* del resultado de las *células complejas*, y así sucesivamente (Figura 3, capas Us2 y Uc2). Es decir, aplicó nuevos *filtros*, en este caso, a las *células complejas* activadas, lo que permitió que la red extrajera características de las imágenes cada vez más complejas. Lindsay explica de la siguiente forma la solución de Fukushima:

> *Simple cells look for patterns; complex cells forgive a slight misplacement of those patterns. Simple, complex; simple, complex. Over and over. Repeating this riff leads to cells that are responsive to all kinds of patterns* (Lindsay, 2022, p. 172).

A través de la repetición de estos dos simples cálculos (convolución y submuestreo) el sistema extraía las características (patrones) de una imagen-*input*, simulando el complejo sistema neuronal de la visión o, al menos, la primera aproximación planteada por Hussel y Wiesel, lo cual fue suficiente para construir lo que sería el primer sistema visual en un computador. Al mostrarle ejemplos de imágenes

del mismo tipo, por ejemplo, diferentes *seises* escritos a mano, el *Neocognitrón* reforzaba las conexiones que definían las características de un *seis*. El modelo aprendía a reconocer una imagen registrando las dependencias y relaciones entre los valores de los puntos que la conforman —puntos que en la actualidad corresponderían a los píxeles de una imagen digital—, lo cual le permitía componer estadísticamente una *representación interna*; es decir, un patrón de conexiones entre valores (líneas, curvas, etc.) que representa una determinada imagen—. Traducido a las redes neuronales contemporáneas:

> *In a photo of an apple, for instance, a red pixel may be surrounded by other red pixels 80% of the time, and so on. In this way also, unusual relations can be combined in more complex graphical features (edges, lines, curves, etc.). Just as an apple has to be recognized from different angles, an actual picture is never memorized, only its statistical dependencies. The statistical graph of dependencies is recorded as a multidimensional internal representation that is then associated to a human-readable output (the word 'apple')* (Pasquinelli, 2017, p. 9).

Los avances en este tipo de redes neuronales para visión artificial se vieron prácticamente paralizados hasta 1998, momento en que nacerían las CNN tal como las conocemos hoy, gracias al científico computacional Yann LeCun, quien realizó una serie de cambios en la estructura del *Neocognitrón*. En primer lugar, LeCun etiquetó cada una de las imágenes de los números escritos a mano con su concepto —el número que representaban las imágenes—, para después incorporar un algoritmo de entrenamiento o aprendizaje, denominado *backpropagation*. Esta técnica debe su nombre a que los cálculos que realiza este algoritmo recorren la red en dirección inversa, desde el output —aquellos valores que se asocian a una determinada etiqueta— hasta el *input* —los valores iniciales de entrada que representan los puntos que conforman la imagen, en este caso el número escrito a mano—. El *backpropagation* utilizará la diferencia entre el resultado producido y el deseado para calcular, capa por capa de la red, cómo afectan al resultado final los diferentes valores de los *filtros* que determinan las conexiones de la red, con el objetivo de actualizarlos y mejorar el porcentaje de acierto. De esta manera, el modelo aprende qué valores son necesarios para identificar imágenes similares a los *inputs* recibidos.

LeCun realizó un entrenamiento de la red con unos 10 000 ejemplos de escritura humana obteniendo buenos resultados, algo que no ocurriría con los modelos de CNN desarrollados a principios del año 2000 para clasificar imágenes fotográficas. Aunque estas redes comenzaran a usar grandes *datasets* de fotografías, con más de 60 000 imágenes, no consiguieron los resultados esperados debido al problema que suponía enfrentarse a imágenes complejas de carácter realista, es decir, fotográficas. Continuando con la propuesta de Fukushima, la solución a este problema fue aumentar el número de capas y conexiones de la arquitectura de la red, lo que permitió extraer patrones más complejos de las imágenes. La primera red en conseguir un 62% de acierto en la clasificación de fotografías fue realizada en 2012 y tenía aproximadamente 80 veces el número de neuronas de la red desarrollada por LeCun. Esto requirió un mayor poder computacional (*hardware*) y muchas más imágenes etiquetadas que permitiesen sacar provecho de todas las conexiones de la red. Este segundo punto fue cubierto gracias a ImageNet, una plataforma con un masivo *dataset* de imágenes editadas y preparadas para el entrenamiento de redes neuronales. En 2012 ImageNet impulsó definitivamente el desarrollo de las actuales CNN, a través de la organización del que fuera durante varios años (2010-2017) el mayor concurso de reconocimiento de imágenes a través de visión artificial (*ImageNet Large Scale Visual Recognition Challenge*).

Con la llegada de las CNN, la influencia de las redes neuronales artificiales en el estudio de la percepción visual se hizo más relevante. Tal como indica Lindsay:

> *The relationship between convolutional neural networks and the brain does not go only one way. Neuroscientists have come to reap rewards from the effort computer scientists put into making models that can solve real visual problems. That's because not only are these large, heavily trained convolutional neural networks good at spotting objects in images, they're also good at predicting how the brain will respond to those same images* (Lindsay, 2022, p. 179).

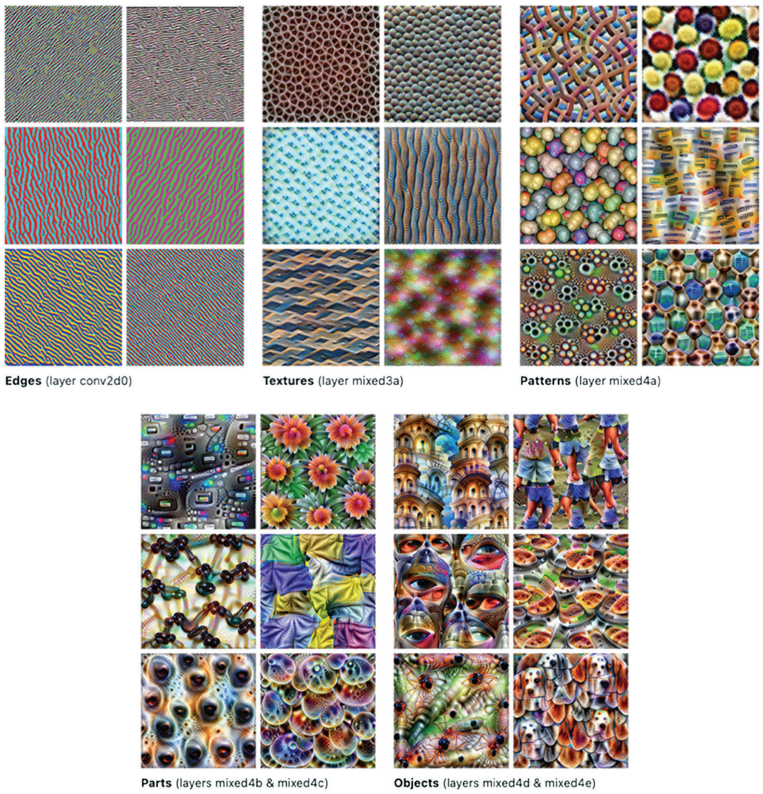
En este sentido, numerosos estudios realizados en neurociencia entre los años 2014 y 2020 establecieron correlaciones entre determinadas capas de las CNN y determinados grupos neuronales implicados en la visión. Entre los métodos computacionales usados en estas investigaciones, destaca la *visualización de características* (Olah et al., 2017), un método que busca hacer visible al ojo humano los cálculos multidimensionales llevados a cabo por las redes CNN durante su proceso de extracción de características de las imágenes, con el objetivo de saber dónde y por qué funcionan bien o mal estas redes. Este método reveló la existencia de patrones visuales preferidos por la CNN que se correlacionan con los encontrados en la neurociencia, como es la detección de líneas en las primeras capas de la red, contornos en las capas posteriores y formas complejas en las capas finales (Figura 5). En concreto, estos métodos se utilizaron para comparar los procesos de las CNN con el comportamiento de la actividad neuronal del sistema visual en monos, mostrándoles para ello las mismas imágenes (Rajalingham et al., 2018). Estas comparaciones demostraron que las neuronas de la última capa de la CNN ayudaban a predecir la actividad de las neuronas ubicadas en la que se considera la última fase del procesamiento visual biológico: la parte inferior del córtex temporal denominada IT, cuyas neuronas responden a objetos concretos, lo que a su vez demostraba que las CNN imitaban la jerarquía visual del cerebro (Lindsay, 2022, p. 181), deconstruyendo la imagen-*input* para volver a reconstruirla en las capas posteriores. En conclusión, se confirmaba que, si las *representaciones internas* no se reconstruyeran a partir de características separadas, si el mundo real se proyectara directamente en el área IT del cerebro, el significado de cada objeto o evento tendría que aprenderse de nuevo, ya que el sistema neuronal implicado no podría realizar nuevas combinaciones de características conocidas (Aleksander, 2002, p. 262).

2. La inducción como limitación del *deep learning* y las CNN

En el apartado anterior he presentado una breve historia de la relación simbiótica que existió entre la neurociencia y la computación, donde las redes neuronales artificiales se inspiraban en patrones biológicos y, posteriormente, los neurocientíficos recurrían a las investigaciones en IA conexionista para identificar el papel computacional de los detalles biológicos. Pero esta relación también trajo consigo una serie de implicaciones discursivas que ayudaron a consolidar un discurso radical, según el cual la suma de los diferentes modelos *deep learning* —entre ellos la visión artificial— permitirá que los futuros sistemas alcancen una cognición como la humana (Reuters, 2024), si es que aún no lo han hecho, tal como llegó a afirmar Elon Musk al decir que los coches Tesla tenían una mente: *I think we may have figured out some aspects of AGI (artificial general intelligence). The car has a mind. Not an enormous mind, but a mind nonetheless* (Musk, 2023).

De acuerdo con Erik J. Larson (2022) y Kate Crawford (2021), este discurso público responde a una mitología de la IA que erosiona la frontera entre lo humano y lo artificial, donde la agencia autónoma

Figura 5: Visualización de las características de unas pocas neuronas activadas en 5 de las capas de una red CNN denominada Inception V1.



Fuente: Olah et al., 2017

de la IA puede reemplazar, e incluso superar, la inteligencia humana. Esta mitología tuvo su inicio con determinados presupuestos sobre la cognición y la inteligencia humana que fueron naturalizados desde la década de los 70 por la ciencia cognitiva *computacionalista*, teoría según la cual la mente es un software y el software de un computador digital podría llegar a ser una mente (Searle, 1984; 2002). Aunque la IA ha cambiado desde la década de los 70 —concibiéndose en la actualidad como una ingeniería y no ya como parte de una ciencia cuyo objetivo era producir una mente artificial—, la narrativa en torno a sus investigaciones parece girar en torno a producir —en lugar de simplemente simular— fenómenos psicológicos como percepción, comprensión, razonamiento, etc. (Preston, 2002, pp. 14-15); de lo contrario, se pregunta Larson (2022), «¿a cuento de qué, por ejemplo, usamos el término "inteligencia artificial" en vez de, quizá, el de "simulación de tareas humanas"?» (p. 41).

Con la llegada del conexionismo y, posteriormente, del *deep learning*, la mitología *computacionalista* se transformó en una *neuromitología* que, de acuerdo con el filósofo y neurocientífico Rymond Tallis (2004, p. 36), es el resultado de una asimilación pasiva y no crítica de los hábitos del lenguaje provenientes de la neurociencia. En el *léxico neuromitológico*, un término como *información* —que en neurofisiológica se refiere a disparos electroquímicos—, se equipara al *conocimiento* que tiene lugar en un nivel superior (psicológico) y a los *datos* con los que opera la IA (símbolos formales y abstractos). Según Tallis:

> *The most important characteristic of these terms is that they have a foot in both camps: they can be applied to machines as well as to human beings and their deployment erodes, or elides, or conjures*

> *away, the barriers between man and machine, between consciousness and mechanism* (Tallis, 2004, p. 34).

Siguiendo el planteamiento de John Preston (2002, p. 33), esta *neuromitología* permite afirmar que una red neuronal artificial configurada y entrenada adecuadamente tiene propiedades psicológicas relevantes, y lo hará simplemente por tener la configuración y el entrenamiento en cuestión. Un ejemplo de ello serían las afirmaciones del ingeniero e investigador Igor Aleksander (2002, pp. 263-264), para quien la capacidad de deconstrucción y reconstrucción que tiene una red neuronal implica que, a diferencia de lo que ocurría con la IA simbólica de los 70 —que no podía *comprender* nada porque solo representaba objetos como símbolos en relación con otros símbolos—, los modelos conexionistas pueden *comprender* las características que extraen de los datos. En el caso de una CNN, las características de las imágenes recibidas le permitirán, siguiendo los argumentos de Aleksander, *imaginar* nuevas representaciones, porque tiene la capacidad de reconstruir apropiadamente objetos que no han sido específicamente memorizados. Según Aleksander (2002, p. 250), esto significa que el sistema opera de forma intencional y *dirigida* a las características de los objetos que percibe —algo que, como veremos a continuación, define a la percepción humana y no a la visión computacional—.

A diferencia del discurso *neuromitológico* en el que se sustentan estas afirmaciones, el neurocientífico y filósofo de la mente John Searle (1984; 2002) defiende un naturalismo biológico por el cual la mente es una propiedad emergente del cerebro. Es decir, independientemente de los avances tecnológicos, la IA nunca podrá *duplicar* los poderes causales de la neurobiología —las conexiones electroquímicas de las neuronas ni lo que causan: el aprendizaje, la consciencia... la mente, en definitiva—, una vez que la IA es un tipo de *máquina* basada en el cálculo y no en la física que determina nuestra realidad causal. La IA solo podrá simular el funcionamiento de las tareas cognitivas a través de cálculo, sin pretender que emerjan propiedades mentales del computador. De acuerdo con Searle (como se citó en Faigenbaum, 2001), hasta que no sepamos los mecanismos (electroquímicos y estructurales) por los cuales las redes neuronales llevan a cabo tareas más complejas, no podremos realizar una simulación causal de su funcionamiento.

En este sentido, aunque el *deep learning* sea descrito por los ingenieros como una tecnología que *aprende de forma autónoma* a través de redes neuronales artificiales que se basan en el funcionamiento de nuestro cerebro (Pachón, 2024; 2025), la base neurobiológica del *deep learning* se reduce a su sistema de conectividad —como hemos visto en la visión artificial, donde se simula el funcionamiento de las *células simples* y *complejas* de la corteza visual primaria—. En este sentido, la mayor fortaleza del *deep learning* proviene, paradójicamente, de aquello que lo limita como simulador neurobiológico: la técnica del *backpropagation*. De acuerdo con Lindsay (2022), las neuronas biológicas sólo pueden conocer la actividad de las neuronas a las que están conectadas, no la actividad de las neuronas a las que se conectan esas otras neuronas, por lo que el *backpropagation* descrito en la sección anterior no se puede considerar lo suficientemente plausible como para ser una aproximación de cómo *aprende* realmente el sistema visual biológico, siendo apenas una solución matemática para que las redes multicapa puedan funcionar, una vez que sería imposible ajustar *manualmente* los valores de los filtros que permiten la conectividad de la red y la consecuente reconstrucción de las imágenes-*input*. Según Lindsay, no podemos simular computacionalmente el funcionamiento del *aprendizaje* de las redes neuronales del cerebro porque aún no sabemos cómo funcionan (p. 79).

De esta forma, sería más apropiado definir el *deep learning* como un conjunto de técnicas computacionales que se limitan a realizar inferencias inductivas, algo que es insuficiente para simular la cognición humana, independientemente de los avances en computación o del descubrimiento de nuevos algoritmos. Ni el *backpropagation*, ni ninguna otra técnica computacional conocida, tiene la capacidad de introducir en la parte formal del sistema la intuición humana, esa curiosidad que nos permite formular nuevas preguntas y establecer conjeturas. Como indica Larson (2022), los diseñadores usarán su intuición, desde fuera del sistema, para introducirle a la IA un sesgo que limita los problemas que debe aprender a resolver (p.42). Además, como señala el filósofo Matteo Pasquinelli (2017, pp. 10-11), el *deep learning* no puede escapar a la ontología categórica en la que opera, lo que le impide plantear nuevos problemas por sí mismo. Es decir, el *deep learning* trabaja dentro de los postulados y categorías humanas que están en los *datasets* de entrenamiento, no puede inventar categorías nuevas.

De acuerdo con Pasquinelli, aunque el *deep learning* realiza un tipo de inferencia, esta será siempre dependiente del humano:

> *Current techniques of Artificial Intelligence are clearly a sophisticated form of pattern recognition rather than intelligence, if intelligence is understood as the discovery and invention of new rules. To be precise in terms of logic, what neural networks calculate is a form of statistical induction. Of course, such an extraordinary form of automated inference can be a precious ally for human creativity and science (...), but it does not represent per se the automation of intelligence qua invention, precisely as it remains within "too human" categories* (Pasquinelli, 2017, p. 9).

La inteligencia que presentamos los humanos en la percepción de nuestra realidad no se limita a los datos y la estadística de la inducción, ya que solo otorgaría conocimiento de algo que sucede regularmente y, además, no puede mostrar las causas de esa repetición en lo observable. La deducción tampoco es suficiente, una vez que si nos encontramos con una excepción que escapa a las proposiciones conocidas no podríamos inferir su causa. Por el contrario, nuestro conocimiento del mundo "depende de la detección sensitiva de la anormalidad o de las excepciones" (Larson, 2022, p. 154); necesitamos de una susceptibilidad a la sorpresa cuyas explicaciones no dependen de generalizaciones ni de expectativas, sino de aquello que Charles S. Peirce denominó *abducción*: una inferencia natural o instintiva que nos permite pasar de la observación de un hecho particular (excepcional) a una hipótesis que parezca probable o verosímil dentro de un mundo de posibilidades infinitas.

La inferencia abductiva nos permite crear las infraestructuras necesarias para dar sentido a la experiencia sensorial del mundo, por lo que podemos decir que la abducción precede a las inducciones y las deducciones. En este sentido, más allá de ser una otra inferencia consciente —realizada a través de la manipulación de proposiciones o símbolos—, la abducción también tiene lugar de forma primaria, a través de la percepción directa de las experiencias del mundo, produciendo un tipo de conjeturas que preceden a cualquier inferencia consciente, como es el caso de la *abducción visual primaria* que veremos a continuación. El filósofo Lorenzo Magnani coloca el siguiente ejemplo de abducción visual:

> *At a certain level of abstraction, visual stimuli, for example, can be viewed as premises, and the outputs of perceptual processing —our knowledge that an obscurely seen face belongs to a friend of ours— in*

> *turn be likened to a conjecture derived from the fact or apparent fact that the best causal account of the presence of those stimuli is the presence of our friend* (Magnani, 2014, p. 178).

Aunque muchos estímulos visuales sean ambiguos —como ocurre con el rostro oscuro del ejemplo anterior—, la percepción visual produce un conocimiento rápido e incontrolado que nos permite explicar lo observado (ese rostro oscuro podría ser nuestro amigo). La perspectiva naturalista de la experiencia visual, como la ofrecida por Searle (2018), nos dice que el mecanismo abductivo emerge, en primera instancia, de la actividad no consciente de las conexiones neuronales, generando un *contenido* que podemos aceptar —algo que sucede normalmente de forma espontánea— o someter a un análisis posterior (Magnani, 2014, p. 175). De esta forma, Magnani (2014, pp. 186-187) establece tres niveles diferenciables en la experiencia visual:

- la *sensación visual*, que el autor define como la formación de imágenes retinianas que aún no son útiles a nivel cognitivo;

- la *percepción*, que corresponde con la actividad de las conexiones neuronales que transforman la *sensación visual* en una *representación estructurada* —este nivel *perceptivo* será el que simula la CNN a través de la deconstrucción y reconstrucción de las imágenes-*input*—;

- y, por último, el nivel de la *observación*, que consiste en procesos posteriores que implican *estados cognitivos elevados* —como la creencia, las expectativas, la memoria, etc.—, lo cual posibilita, entre otras inferencias perceptivas, el reconocimiento o la identificación de los objetos y eventos *percibidos*, como era el caso del rostro oscuro del ejemplo anterior —algo que las CNN no pueden simular, siendo que apenas ofrecen una solución técnica que le permite realizar un método inductivo enumerativo, comportándose *como si* tuviera la capacidad de *observar*—.

El nivel de la *observación* implica que la percepción visual no funciona aislada de otros modos de percepción o fuentes de experiencia. Pero esto no quiere decir que no exista una *abducción primaria* en el *nivel perceptivo*. Por ejemplo, la frase *estoy viendo un árbol* es abstracta —un conjunto de palabras (símbolos)—, por lo que pertenece al nivel elevado de la *observación*; pero el árbol que yo percibo es algo concreto, una señal que *apunta* hacia una característica del mundo, algo que hago inteligible a través de dicha frase. Es decir, según Magnani (2014, p. 186), hay una abducción en el *nivel perceptivo* que genera *representaciones primarias* con las que posteriormente opera el nivel de la *observación*, de tal manera que las percepciones adquieren significados. Según Magnani:

> *Higher cognitive states affect the product of visual modules only after the visual modules '[...] have produced their product, by selecting, acting like filters, which output will be accepted for further processing' [Raftopoulos 2001, p. 434], for instance by selecting through attention, imagery, and semantic processing, which aspects of the retinal input are relevant, activating the appropriate neurons. (...) I consider these processes essentially abductive* (Magnani, 2014, p. 187).

Tras leer esta explicación, podríamos preguntarnos si esto no es exactamente lo que hacen las CNN: usar unos *filtros* que permiten activar las neuronas (valores numéricos en una función) para extraer las características *relevantes* de las imágenes de entrada, permitiendo operar con ellas posteriormente —es

decir, clasificar las imágenes en categorías—. Si aceptamos esta comparativa podríamos, en el mejor de los casos, considerar que la CNN simula un tipo de *abducción débil* (Pasquinelli, 2017), una vez que, si bien extrae características *relevantes* que le permiten clasificar nuevas imágenes, lo hace a través de una inducción estadística basada en datos previamente etiquetados por los humanos, donde la *relevancia* atiende exclusivamente al contexto del *dataset* de entrenamiento. Los procesos de una CNN son, como todo cálculo computacional, *relativos al observador*[1], y los *outputs* que genera son apenas representaciones abstractas (conjuntos de unos y ceros) que responden a procesos puramente sintácticos y, por tanto, no se refieren a nada concreto del mundo. Por su parte, las *representaciones internas* generadas por la abducción visual biológica se refieren a configuraciones particulares de redes neuronales que, como indica Searle (2018, p. 44), se *dirigen* a las características del mundo.

Conclusión

Las teorías de la abducción nos muestran que las inferencias que tienen lugar durante la actividad visual de los seres humanos responden a procesos semánticos que se *dirigen* y se *ajustan* a las características concretas del mundo. Por su parte, el *deep learning* y las CNN no tienen semántica, no conocen nada del mundo; tan solo realizan procesos sintácticos a través de operaciones lógicas con *datos* (números). Si bien el conexionismo, el *deep learning* y las CNN fueron fundamentales para desarrollar modelos de simulación de tareas cognitivas básicas, ello no implica que, por *sobreestimular computacionalmente* una pequeña parte del funcionamiento neuronal implicado en la percepción visual, la *máquina* adquiera la capacidad de *observar*.

Las CNN tratan las características del mundo como *datos* analizables —donde los *datos* se refieren a los píxeles (ceros y unos) de las imágenes digitales previamente realizadas, seleccionadas y etiquetadas por los humanos—. A diferencia de lo que nos dice la *neuromitología*, aprender no es una cuestión de retirar del ambiente *datos* —*información*, según la crítica de Tallis (2004, p. 35)— que satisfagan las condiciones de *input* de un determinado módulo cerebral. Por el contrario, tal como indica el antropólogo Tim Ingold (2001, p. 130), aprender es formar, dentro del ambiente del mundo, las conexiones neurológicas necesarias que, junto a los *estados cognitivos elevados* que indicaba Magnani (2014) —como la creencia, el deseo o la memoria—, dan lugar al nivel de la *observación*.

Referencias bibliográficas

Aleksander, I. (2002). Neural Depictions of 'World' and 'Self': Bringing Computational Understanding to the Chinese Room. En Preston J. y Bishop M. (eds.), *Views into the Chinese Room: new essays on Searle and artificial intelligence*. (pp. 250-268). Oxford University Press.

Altman, Sam. (2024). *The Intelligence Age*. Última modificación 23 de septiembre. https://ia.samaltman.com

Cuartas, J. (30 de enero de 2021). *El concepto de la convolución en gráficos, para comprender las Convolutional Neural Networks (CNN) o redes convolucionadas*. Medium. https://josecuartas.medium.com/el-concepto-de-la-convolución-en-gráficos-para-comprender-las-convolutional-neural-networks-cnn-519d2eee009c#:~:text=sensor%20radar

[1] Recordemos que somos nosotros los humanos, desde fuera del sistema, quienes establecemos las relaciones entre determinados patrones y su relación con el mundo, como muestra el etiquetado del *dataset* con el que las redes neuronales artificiales aprenden.

Crawford, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

Faigenbaum, G. (2001). *Conversaciones con John Searle*. LibrosEnRed. Kindl e-book.

Ingold, T. (2001). From the Transmission of Representations to the Education of Attention. En Whitehouse, H. (ed.), *The Debated Mind. Evolutionary Psychology Versus Ethnography*. (pp. 113 – 153). Berg.

Kaput, M. (5 de marzo de 2024). *Sam Altman Says AI Will Handle '95%' of Marketing Work Done by Agencies and Creatives.* Marketing Artificial Intelligence Institute. https://www.marketingaiinstitute.com/blog/sam-altman-ai-agi-marketing.

Larson, E.J. (2022). *El mito de la Inteligencia Artificial: por qué las máquinas no pueden pensar como nosotros lo hacemos*. Shackleton Books, S.L.

Lindsay, G. (2022). *Models of the Mind: How physics, engineering and mathematics have shaped our understanding of the brain*. Bloomsbury.

Magnani, L. (2014). Understanding abduction. En Magnani, L. (ed.), Model-Based Reasoning in Science and Technology. Theoretical and Cognitive Issues. *Studies in Applied Philosophy, Epistemology and Rational Ethics Vol. 8*, 173-206. Springer-Verlag.

Musk, Elon [@elonmusk]. (7 de agosto 2023). *I think we may have figured out some aspects of AGI. The car has a mind. Not an enormous mind, but a mind nonetheless*. X (Twitter). https://x.com/elonmusk/status/1688476506295586816

Negrón, Juan José. (2020). *Inspiración Biológica de las Redes Neuronales* Convolucionales. Medium. Última modificación 18 de septiembre. https://medium.com/soldai/inspiración-biológica-de-las-redes-neuronales-convolucionales-c686f74b4723

Olah, C., Mordvintsev, A., y Schubert, L. (7 de noviembre de 2017). *Feature Visualization. How neural networks build up their understanding of images.* Distill. https://distill.pub/2017/feature-visualization/

Pachón, A. (2024). Epistemic interfaces of visualization and interpretation: a possible resistance to the myth of autonomous agency in AI. *Revista de Comunicação e Linguagens, 60*, 91-113.

Pachón, A. (2025). The Mythical Speech of Artificial Intelligence: The Imaginary of an Autonomous Agency. En VandenBroek, A. K., Koycheva, L. y Artz, M (eds.), *Anthopology of AI*, 65-82. Routledge.

Pasquinelli, M. (2017). Machines that Morph Logic: Neural Networks and the Distorted Automation of Intelligence as Statistical Inference. *Site 1: Logic Gate, the Politics of the Artifactual Mind*. Glass Bead. https://www.glass-bead.org/article/machines-that-morph-logic/?lang=enview

Preston, J. (2002). Introduction. En Preston J. y Bishop M. (eds.), *Views into the Chinese Room: new essays on Searle and artificial intelligence*. (pp. 1-50). Oxford University Press.

Raftopoulos, A. (2001). Is perception informationally encapsulated? The issue of theory-ladenness of perception. *Cognitive Sciencies, 25*, 423–451.

Rajalingham, R., Issa E. B., Bashivan, P., Kar, K., Schmidt, K., y DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *Journal of Neuroscience, 38*, 7255–7269.

Reuters. (8 de abril de 2024). *Tesla's Musk predicts AI will be smarter than the smartest human next year.* Reuters. https://www.reuters.com/technology/teslas-musk-predicts-ai-will-be-smarter-than-smartest-human-next-year-2024-04-08/

Searle, J. R. (1984). *Minds, Brains and Science*. Harvard University Press.

Searle, J. R. (2002). Twenty-One Years in the Chinese Room. En Preston J. y Bishop M. (eds.), *Views into the Chinese Room: new essays on Searle and artificial intelligence*. (pp. 51-69). Oxford University Press.

Searle, J. R. (2018). *Ver las cosas tal como son. Una teoría de la percepción*. Oxford University Press.
Seisdedos, I. (16 de mayo de 2023). Sam Altman (ChatGPT), en el Capitolio: 'Si la inteligencia artificial sale mal, puede salir muy mal'. *El País*. https://elpais.com/tecnologia/2023-05-16/sam-altman-chat-gpt-en-el-capitolio-si-la-inteligencia-artificial-sale-mal-puede-salir-muy-mal.html
Tallis, R. (2004). *Why the Mind Is Not a Computer. A pocket Lexicon of Neuromythology*. Imprint Academic.

BIO

Doctorado en Antropología Social y Cultural (Universidad de Coimbra), con una beca de la Fundação para a Ciência y a Tecnologia (FCT, Portugal), máster en Antropología Social y Cultural (Universidad de Coimbra), magíster en Teoría y Práctica del Arte Contemporáneo (Universidad Complutense) y licenciado en Bellas Artes (Universidad Complutense de Madrid). Su trayectoria académica y profesional se desarrolla de forma interdisciplinar, entre el arte y la antropología. En 2019 recibió una Beca Leonardo de la Fundación BBVA (España), desarrollando un proyecto que sirvió como punto de partida para su investigación doctoral, en la cual cruza la etnografía colaborativa con la práctica artística para desarrollar una antropología de la Inteligencia Artificial. Su obra artística forma parte de importantes colecciones, como el Museo Nacional Reina Sofía (Madrid), el CA2M (Madrid) y el MEIAC (Badajoz). info@andrespachon.com