

## DIÁLOGOS CIENCIA Y DERECHO

# “ABC de la Inteligencia Artificial y de la Neurociencia”

Resumen de la sesión de 1 de octubre de 2018

**Presentó la sesión: Antonio Garrigues**, Presidente de la Fundación Garrigues

**Ponentes: Francisco Rubia Vila**, Catedrático emérito de las Universidades Complutense y Ludwig Maximilian de Munich y **Ramón López de Mantaras**, Profesor de Investigación de la CSIC y Fundador y Director del Instituto de Investigación en Inteligencia Artificial del CSIC. Miembro del Consejo Académico de FIDE

**Moderador: Pedro García Barreno**, Doctor en Medicina. Catedrático emérito de la Universidad Complutense. Miembro del Consejo Académico de Fide

“**El cerebro nos engaña**”, dice el profesor **Francisco J. Rubia**, por ser una máquina de confabulaciones y fantasías, generadora de música, arte, mitos, religión y creatividad, actividades ligadas al sistema emocional. La mente, como fruto de la actividad cerebral es producto de la evolución y si se adapta al entorno, asegurándose así su pervivencia, no es en él únicamente donde el hombre adquiere lo que mentalmente es, sino que ya al nacer posee predisposiciones genéticas que irá desarrollando a medida que encuentre el ambiente propicio; de ahí la importancia de la cultura. Los pensamientos, místico y lógico, son dos formas con sustratos distintos y, pese a todos los cambios habidos, perduran en el hombre moderno, lo mismo que las estructuras cerebrales que los sustentan.

Por su parte, el profesor **Ramón López de Mántaras** se pregunta: “**¿Es posible construir máquinas inteligentes? ¿Es el cerebro una máquina?**” Estas dos preguntas han sido la obsesión de grandes pensadores durante siglos. Pero con el desarrollo de la inteligencia artificial (IA), ambas cuestiones se han acercado e incluso se han unificado pues utilizan los mismos conceptos, técnicas y experimentos en los intentos de diseñar máquinas inteligentes y en investigar la naturaleza de la mente. El objetivo último de la inteligencia artificial —lograr que una máquina tenga una inteligencia de tipo general similar a la humana— es de los más ambiciosos que se ha planteado la ciencia. Por su dificultad, es comparable a otros grandes objetivos científicos como explicar el origen de la vida, el del universo o conocer la estructura de la materia. Actualmente, todavía sabemos poco acerca del cerebro; sin embargo, estamos siguiendo un camino que pasa por considerarlo un sistema computacional y hemos empezado a explorar el espacio de posibles modelos computacionales que permitan emular su funcionamiento.

La sesión ABC de la neurociencia e inteligencia artificial” —un diálogo abierto entre ambos—, pretendió ofrecer las bases, compartidas unas y distantes otras, de la neurofisiología y la IA.

Abrió la sesión **Antonio Garrigues**, Presidente de la Fundación Garrigues, quién resaltó el impacto bidireccional entre la ciencia y la técnica y el derecho: Neurociencia y Ley, Robótica y Ley, *Big data* y Ley... Remachó la necesidad del abordaje a los problemas mediante la convergencia de saberes; también la ineludible obligación de reformar un currículo formativo en educación primaria y secundaria que apunta en dirección opuesta.

El **Prof. Francisco Rubia Vila** se esforzó en presentar una panorámica general y básica de la estructura y función del sistema nervioso, haciendo hincapié en aquellas cuestiones solapantes con la Inteligencia Artificial (IA). Inició su intervención refiriéndose a la **estricta vinculación de la presencia de un sistema nervioso a la movilidad del organismo**. Aquellos que permanecen anclados desde su inicio carecen de tal sistema, y otros que inicialmente buscan sustento en el que vivir, desarrollan un sistema nervioso rudimentario que desaparece tras su establecimiento definitivo.

Cuatro partes anatómicamente bien definidas –cerebro, cerebelo, puente y médula espinal- forman el sistema nervioso del individuo adulto, formado mediante el desarrollo de una serie de vesículas primitivas o embrionarias. La médula espinal se localiza en el canal vertebral y es responsable de la conducción de estímulos desde las estructuras centrales a la periferia (diversos órganos y sistemas) y viceversa. El puente es una estructura a modo de unión entre cerebro y médula, que alberga importantes centros vitales (respiración, cardio-circulatorio). El cerebelo es una estructura esencialmente involucrada en el control del movimiento, incluido el lenguaje. **El cerebro es la estructura dominante que adquirió durante la evolución el máximo volumen en la especie humana.**

Desde el punto de vista evolutivo, Paul MacLean considera el cerebro a modo de una estructura “triúnica” formada por la agregación sucesiva de los cerebros reptiliano, límbico y cortical. El primero asumiría las funciones básicas para sobrevivir; el segundo, sería responsable de las relaciones interindividuales –aquí se refirió a que si bien un perro no “entiende” las palabras de su amo, si sabe interpretar los sentimientos de éste-, y el córtex asume las funciones cognitivas cada vez más refinadas y complejas de la evolución de las especies.

El peso promedio del cerebro humano es, aproximadamente, 1450 g; su volumen, 1200 ml; representa el 2% del peso corporal, aunque recibe el 15% del gasto cardiaco y consume el 20% del total de oxígeno y el 25% de la glucosa. **Todo ello proporciona una potencia de 20 vatios.**

El cerebro se estructura en dos hemisferios, cada uno contiene un lóbulo frontal, parietal, temporal y occipital. Los hemisferios están unidos por un sistema de conexiones que, en conjunto, se denomina cuerpo calloso con 200 millones de fibras. La parte externa de los lóbulos cerebrales es irregular en cuanto está surcada por numerosos pliegues que incrementan la superficie. La parte más exterior se denomina

corteza o sustancia gris y está formada, principalmente por células denominadas neuronas; la parte más profunda se denomina sustancia blanca, dominada por fibras de conexión.

El número de neuronas, pleomorfas anatómica y funcionalmente, asciende a cien mil millones, aproximadamente el número de estrellas en nuestra galaxia, y se conectan entre sí; cada neurona con, aproximadamente, mil congéneres. Ello produce una complejidad de **400 billones de conexiones**; tal vez la estructura más compleja de la evolución. Un pequeño universo de 1400g que opera, aproximadamente, a una velocidad de 1 petaflop (FLOPs: operaciones por segundo), esto es,  $10^{15}$  operaciones  $sc^{-1}$ .

Las neuronas se ordenan por capas y columnas; cada una de ellas con estructuras particulares en cada zona cerebral. Se admite que láminas y columnas representan las unidades operativas básicas del cerebro. Las neuronas se ordenan por capas y columnas; cada una de ellas con estructuras particulares en cada zona cerebral. Se admite que láminas y columnas representan las unidades operativas básicas del cerebro. En términos generales, una neurona está formada por un cuerpo y dos clases de excrescencias. Unas, **dendritas**, a modo de ramas de un árbol, reciben señales de otras congéneres; otra, única, denominada **axón**, sirve de conducción de las señales neuronales a hacia sus dianas. Las conexiones se establecen a través de estructuras denominadas **sinapsis**. En condiciones de reposo la neurona mantiene un equilibrio eléctrico a través de su membrana; cuando se activa, reacciona mediante un mecanismo tipo todo-o-nada alterando el potencial de reposo de la membrana que se transforma en un potencial de acción responsable de cambios de permeabilidad en la hendidura sináptica; ello provoca la liberación de mediadores químicos prefabricados y almacenados (neurotransmisores) que interaccionan con la membrana postsináptica de la siguiente neurona de la cadena, provocando su activación o inhibición.

Las diferentes partes del cerebro, cuyas fronteras están bien señaladas por determinadas cisuras, están especializadas. Los lóbulos occipitales involucrados en la visión; parietales en el procesamiento sensorial; los temporales en el reconocimiento de formas y la memoria, y los frontales, los más desarrollados en la especie humana (proceso de encefalización): procesamiento de funciones cognitivas y control voluntario de los músculos (corteza motora, justo por delante de la corteza sensorial parietal. En ambas puede representarse un mapa de la topografía cerebral, cuya extensión es proporcional a la "importancia" en receptores de la estructura correspondiente).

Aparte de la estructura neuronal de la corteza existen **conglomerados neuronales en el centro del cerebro (núcleos de la base) con importantes funciones y almacenamiento de las habilidades motoras**. Por su parte, la parte central del cerebro está formada por un sistema de **ventrículos y conductos** que se prolongan por la médula espinal; por él circula el líquido cefalorraquídeo, fluido por el que circulan nutrientes y se retira material de desecho.

Aparte de la división anatómica lobular, las diferentes funciones están bien delimitadas, aunque sigue en discusión el peso relativo entre la “individualidad” estructura-función y la funcionalidad global cerebral. En principio, diferentes lesiones cerebrales, por accidente o enfermedad, y, posteriormente, mediante las diferentes tecnologías diagnósticas, desde la radiología convencional a la tractografía, pasando por la tomografía computarizada (TC), tomografía por emisión de positrones (PET), imagen por resonancia (IRM), electroencefalografía (EEG) y magnetoencefalografía (MEG), se han podido “localizar” diferentes funciones “superiores”; tal es el caso de sistema del habla y del lenguaje. A nivel celular, técnicas singulares han identificado **neuronas especializadas en detectar nuestra localización y posición**, una especie de GPS cerebral; o reconocer e interiorizar lo que hacen los demás (neuronas espejo).

Refiriéndose a su libro “El Cerebro Nos Engaña” (Editorial: Temas de Hoy, Madrid, 2000) el Prof. Rubia concluye con el problema mente-cerebro. **Conocemos la estructura y el resultado funcional, pero existe un “gap” entre ambos**. Se refiere al **nematodo *Caenorhabditis elegans***; tiene 302 neuronas, se conocen todas y cada una de sus conexiones, pero se desconoce la “esencia” del funcionamiento de un sistema nervioso tan simple. **El reto es desentrañar cómo emerge la consciencia de ese sistema hipercomplejo que es el cerebro**.

A continuación, el **Prof. Ramón López de Mántaras Badia** inició su intervención sobre inteligencia artificial (IA), con las preguntas: ¿Es posible construir máquinas inteligentes? ¿Es el cerebro una máquina? Tal es el comienzo de la «Introducción» de su reciente libro “*Inteligencia artificial*” - publicado en colaboración con Pedro Meseguer, en la colección ¿Qué sabemos de?, editada por el CSIC-Catarata, Madrid 2017 – y que sirvió de guion a lo largo de su intervención. En la Introducción puede leerse: Estas dos preguntas han sido la obsesión de grandes pensadores durante siglos. Pero con el desarrollo de la inteligencia artificial, **ambas cuestiones se han acercado e incluso se han unificado pues utilizan los mismos conceptos, técnicas y experimentos en los intentos de diseñar máquinas inteligentes y en investigar la naturaleza de la mente**. El objetivo último de la inteligencia artificial —lograr que una

máquina tenga una inteligencia de tipo general similar a la humana— es de los más ambiciosos que se ha planteado la ciencia.

Por su dificultad, es comparable a otros grandes objetivos científicos como explicar el origen de la vida, el del universo o conocer la estructura de la materia. Actualmente, todavía **sabemos poco acerca del cerebro**; sin embargo, **estamos siguiendo un camino que pasa por considerarlo un sistema computacional y hemos empezado a explorar el espacio de posibles modelos computacionales que permitan emular su funcionamiento**. El objetivo último de la IA – lograr que una máquina tenga una inteligencia de tipo *general* similar a la humana – es de los más ambiciosos que se ha planteado la ciencia. Por su dificultad, es comparable a otros grandes objetivos científicos como explicar el origen del universo, la estructura de la materia o el origen de la vida.

**La inteligencia no es una característica exclusiva de los humanos.** En la naturaleza existen muchos animales que exhiben un comportamiento inteligente, en el sentido de que planifican, son capaces de prever las consecuencias de sus acciones y emplean útiles o herramientas para conseguir sus propósitos. Algunos animales tienen también capacidades, aunque muy limitadas, para procesar el lenguaje. Por este motivo es más correcto hablar de **inteligencias** que de inteligencia, **y no sería absurdo pensar que la IA pueda llegar a constituir un nuevo tipo de inteligencia, aunque distinta a la de los animales y humanos.**

Centrándose en la inteligencia humana, que es el referente principal de la IA, es importante distinguir entre **IA débil e IA fuerte**, dos visiones que se corresponden, respectivamente con los dos siguientes intentos de definición: **1. La IA es la ciencia e ingeniería que permite diseñar y programar ordenadores de forma que realicen tareas que requieren inteligencia. 2. La IA es la ciencia e ingeniería que permitirá replicar la inteligencia humana mediante máquinas.**

Es importante el matiz que debería tratarse de inteligencia de tipo general y no especializada, ya que la **inteligencia de los seres humanos es de tipo general**. Quién introdujo esta distinción entre IA débil e IA fuerte fue John Searle, en 1980, que provocó mucha polémica. La IA fuerte implicaría que un ordenador convenientemente programado no simula una mente, sino que es una mente y por consiguiente debería ser capaz de pensar igual que un ser humano. Searle, trata de demostrar que la IA fuerte es imposible. La IA débil consistiría, según Searle, en construir programas que ayudan al ser humano en sus actividades mentales en vez de

duplicarlas. La capacidad de los ordenadores para realizar tareas específicas mejor que las personas ya se ha demostrado. Todos los avances logrados hasta ahora en el campo de la IA son manifestaciones de la IA débil. Concluyó Ramón su introducción refiriéndose a **los principales modelos de IA: simbólico, conexionista, evolutivo y corpóreo. Esta última aproximación ha ido ganando terreno en la IA: el cuerpo da forma a la inteligencia; sin cuerpo no puede haber inteligencia de tipo general.**

Respecto a los orígenes de la IA no pueden entenderse sin hablar de Alan M. Turing (1912-1954); un precursor y visionario de la IA. El nacimiento oficial de la IA se fija en un encuentro científico en el Dartmouth College, en New Hampshire, en 1956, aunque un año antes se había celebrado una sesión dedicada a *máquinas que aprenden* en Los Ángeles, California.

Una clase de problemas estudiados desde los primeros años de la IA son los juegos de tablero y, en particular, el ajedrez, juego al que se han referido como la *Drosophila* de la IA. En 1962 se desarrolló el primer programa capaz de jugar a un juego de tablero a un nivel avanzado; se trataba del juego de *checkers*, similar a las damas. Dado que hay  $5 \times 10^{20}$  posiciones posibles, era imposible evaluar todo el árbol de búsqueda. Mediante aprendizaje, la calidad de la función heurística – conjunto de técnicas o métodos para resolver un problema. La heurística es vista como el arte de inventar por parte de los seres humanos, con la intención de procurar estrategias, métodos, criterios, que permitan resolver problemas a través de la creatividad, pensamiento divergente o lateral. También, se afirma que la heurística se basa en la experiencia propia del individuo, y en la de otros para encontrar la solución más viable al problema – mejoraba. Otra clase de dificultades que motivaron los primeros programas capaces de aprender, era los de reconocimiento y clasificación de patrones. El paso siguiente es entender el significado de los términos y frases que forman parte de dichos caracteres. Este es el objetivo de lo que en IA se conoce por «procesamiento del lenguaje natural».

Los primeros grandes éxitos de la IA, concretamente en los temas de representación del conocimiento, búsqueda heurística, planificación, sistemas expertos, aprendizaje automático, visión por computador, robots móviles y procesamiento del lenguaje escrito y hablado. A principios de los años setenta empezaron a aparecer voces críticas con la IA, algunas de carácter filosófico y otras de carácter ético, alrededor de la distinción entre mente y máquina (James Lighthill, John Searle, Hubert Dreyfus, Roger Penrose o Joseph Weizenbaum – uno de los más activos en contra el uso socialmente irresponsable de la IA). También en esa época,

los investigadores en IA se dieron cuenta de que existían serias dificultades de carácter técnico con el problema de la explosión combinatoria en los sistemas de IA. Todo ello contribuyó a lo que se conoce como el *invierno* de la IA, que se prolongó hasta mediados de los ochenta cuando se produjo un cambio de paradigma: se abandonó el objetivo de lograr una IA de tipo general para centrarse en la IA especializada.

**La IA resurgió con la aparición de nuevas e importantes aproximaciones en todas sus áreas, desde la representación de conocimientos y el razonamiento hasta los sistemas multiagente, pasando por el aprendizaje, la visión y la robótica.** La inclusión de la **lógica borrosa** supuso una verdadera revolución y palanca de avance y, en su entorno, **las lógicas de la descripción o el razonamiento basado en restricciones. Por otro lado, las redes bayesiana, la planificación y, sobre todo, el aprendizaje automático en sus diferentes variedades.**

Con todo ello, los grandes éxitos de la IA han sido muy mediáticos: juegos como Deep Blue o AlphaGo, lenguajes como Watson, Mastor o Siri, vehículos autónomos, robots en Marte, videojuegos...

El último capítulo del libro de Ramón comentado, el sexto, lleva por título "Futuro" y en su exposición argumentó que, **diseñar sistemas que tengan las capacidades más complicadas de alcanzar (aquellas que requieren interactuar con entornos no restringidos ni previamente preparados), requiere integrar desarrollos en todas las áreas de la IA.** En particular, necesitamos lenguajes de representación de conocimientos que codifiquen información acerca de muchos tipos distintos de objetos, situaciones, acciones, etc., así como de sus propiedades y de las relaciones entre ellos, **en particular relaciones causa-efecto.** También necesitamos nuevos algoritmos que, en base a estas representaciones, puedan, de forma robusta y eficiente, **resolver problemas y responder preguntas sobre prácticamente cualquier tema.**

Finalmente, dado que necesitarán adquirir un número prácticamente ilimitado de conocimientos, estos sistemas deberán ser capaces de aprender de forma continua a lo largo de toda su existencia. **En definitiva, es imprescindible diseñar sistemas que integren percepción, representación, razonamiento, acción y aprendizaje. Éste es un problema muy importante en IA ya que todavía no sabemos cómo integrar todos estos componentes de la inteligencia.** Necesitamos arquitecturas cognitivas que integren a estos componentes de forma adecuada. **Los sistemas integrados son un**



**paso previo fundamental para conseguir algún día inteligencias artificiales de tipo general.**

Entre las actividades futuras, añadió que los temas de investigación más importantes pasarán por **sistemas híbridos que combinen las ventajas que poseen los sistemas capaces de razonar en base a conocimientos y uso de la memoria y las ventajas de la IA basada en análisis de cantidades masivas de datos, es decir en lo que se conoce por aprendizaje profundo.** Actualmente una importante limitación de los sistemas de aprendizaje profundo es lo que se conoce como **“olvido catastrófico”** lo cual significa que si una vez han sido entrenados para llevar a cabo una tarea (por ejemplo jugar a Go), si a continuación los entrenamos para llevar a cabo otra tarea distinta (por ejemplo distinguir entre imágenes de perros y de gatos) olvidan completamente la tarea anteriormente aprendida (en este caso jugar a Go). **Esta limitación es una prueba contundente de que en realidad estos sistemas no aprenden nada, por lo menos en el sentido humano de aprender.** Otra importante limitación de estos sistemas es que son **“cajas negras” sin capacidad explicativa**, por ello un objetivo interesante de investigación será como dotar de capacidad explicativa a los sistemas de aprendizaje profundo incorporando módulos que permitan explicar cómo se ha llegado a los resultados y conclusiones propuestas, ya que la capacidad de explicación es una característica irrenunciable en cualquier sistema inteligente.

**También es necesario desarrollar nuevos algoritmos de aprendizaje que no requieran enormes cantidades de datos para ser entrenados así como hardware mucho más eficiente en consumo energético para implementarlos ya que el consumo de energía podría acabar siendo una de las principales barreras al desarrollo de la IA.** En comparación el cerebro es varios órdenes de magnitud más eficiente que el hardware actual necesario para implementar los algoritmos de IA más sofisticados. **Una posible vía a explorar es la ingeniería o computación neuromórfica basada en memristores “memoria”- “resistor”:** fue el cuarto elemento fundamental de los circuitos que simulan, en términos generales, una neurona artificial.

Otras técnicas más clásicas de IA que seguirán siendo objeto de investigación extensiva son los **sistemas multiagente**, la planificación de acciones, el razonamiento basado en la experiencia, la visión artificial, la comunicación multimodal persona-máquina, la robótica humanoide y en particular las nuevas tendencias en robótica del desarrollo, que puede ser la clave para dotar a las máquinas de sentido común y en particular aprender la relación entre sus acciones y los efectos que éstas producen en el entorno. También veremos progresos significativos gracias a las **aproximaciones**

**bio-miméticas para reproducir en máquinas el comportamiento de animales.** No se trata únicamente de reproducir el comportamiento de un animal sino de comprender como funciona el cerebro que produce dicho comportamiento. Se trata de construir y programar circuitos electrónicos que reproduzca la actividad cerebral que genera este comportamiento.

Algunos biólogos están interesados en los intentos de fabricar un cerebro artificial lo más complejo posible, porque consideran que es una manera de comprender mejor el órgano y los ingenieros buscan información biológica para hacer diseños más eficaces. **Mediante la biología molecular y los recientes avances en opto-genética será posible identificar que genes y que neuronas juegan un papel clave en las distintas actividades cognitivas.**

En cuanto a las **aplicaciones**, algunas de las más importantes seguirán siendo aquellas relacionadas con la web, los video-juegos, los asistentes personales y los robots autónomos (en particular vehículos autónomos, robots sociales, robots para la exploración de planetas, etc). Las aplicaciones al medio ambiente y ahorro energético también serán importantes, así como las aplicaciones a la economía y la sociología.

Por último, las aplicaciones de la IA Inteligencia Artificial al arte (artes visuales, música, danza, narrativa) cambiarán de forma importante la naturaleza del proceso creativo. Los ordenadores ya no son solamente herramientas de ayuda a la creación, los ordenadores empiezan a ser agentes creativos. Ello ha dado lugar a una nueva y muy prometedora área de aplicación de la Inteligencia Artificial denominada **Creatividad Computacional** que ya ha producido resultados muy interesantes en ajedrez, música, artes plásticas y narrativa, entre otras actividades creativas.