

# ***Diálogos con la ciencia:***

**Inteligencia artificial: Progresos y desafíos**

**Resumen de la sesión del 21 de marzo de 2016**

**Presentó la sesión: Pedro García Barreno.** Doctor en Medicina. Catedrático emérito de la Universidad Complutense de Madrid.

**Ponente: Ramón López de Mantaras Badía.** Profesor de Investigación del Consejo Superior de Investigaciones Científicas (CSIC). Fundador y Director del Instituto de Investigación en Inteligencia Artificial del CSIC.

### Resumen elaborado por Wanda Cazalla, abogada.

El pasado 21 de abril, Ramón López de Mantaras compartió con los asistentes a esta sesión de "Diálogos con la Ciencia" los **principales progresos y desafíos que presenta la Inteligencia Artificial en la actualidad.**

Comenzó señalando que, posiblemente, la **lección más importante** que hemos aprendido a lo largo de los 60 años de historia de la Inteligencia Artificial (IA) es que lo que parecía más difícil de lograr (como por ejemplo diagnosticar enfermedades o jugar al ajedrez o a Go a un nivel comparable o superior a los mejores jugadores humanos) ha resultado relativamente fácil y, sin embargo, aquello que hace 50 años parecía más sencillo (reconocer objetos visualmente o comprender el lenguaje) todavía no se ha logrado.

La exposición abordó la razón de esta aparente contradicción, haciendo un breve repaso del desarrollo de la IA y enfatizando la enorme dificultad de dotar de sentido común a las máquinas, condición indispensable para lograr IA de tipo general.

En general, existen dos aproximaciones a la IA, y se da cierta confusión entre ellas. La denominada **IA fuerte** ("strong AI") hace referencia a las bases de la IA sentadas por sus fundadores en el verano de 1956 en el Dartmouth College (EEUU). Bajo esta aproximación, la IA persigue replicar o incluso superar la inteligencia humana mediante máquinas.

A lo largo de las dos décadas siguientes a la "fundación" de la IA, se generaron grandes expectativas sobre sus logros, que sin embargo se vieron en gran parte frustradas en distintas áreas de investigación, lo que llevó en los años 80 a un importante recorte de la financiación de la IA a nivel internacional ("AI winter"). Uno de los intentos fallidos en el área del procesamiento del lenguaje fue la construcción de un traductor automático inglés-ruso, que sin embargo no dio los resultados esperados.

Lo anterior provocó un importante cambio de enfoque que llevó a lo que se conoce como **IA débil** ("weak AI"). Desde una perspectiva mucho más práctica, esta aproximación a la IA persigue el diseño de máquinas con capacidad para realizar tareas que requieren inteligencia.

Existen **importantes diferencias** entre las dos aproximaciones. Así, por ejemplo, a diferencia de los ordenadores construidos hasta la fecha, un humano que está especializado en un área concreta del conocimiento (un médico cirujano) es capaz de dar un diagnóstico generalista, e incluso de hablar de economía; a día de hoy esto no es posible con las máquinas. Por otra parte, por su complejidad el planteamiento de la IA fuerte es una de las grandes preguntas de la Ciencia (junto con otras como el origen de la vida, el origen del universo o la estructura de la materia, etc.), que quizá no se resuelva nunca, pero que motiva la investigación. La hipótesis del sistema de símbolos físicos defiende que el único requerimiento para que exista IA es que LA inteligencia ES INDEPENDIENTE del sustrato físico que la soporta (carbono, silicio, ...). No obstante, se trata de una hipótesis, y como tal no está refutada ni verificada.

Por su parte, la IA débil acota tanto el ámbito de actuación de los problemas que resuelve que es capaz de resolverlos eficientemente. Bajo esta aproximación, todas las IA son débiles. La IA es omnipresente, y existen múltiples ejemplos en nuestro día a día: en logística, transportes, diagnósticos, tomas de decisiones, ... Cualquier sector tiene hoy en día algún elemento de IA,

incluidos la música y el deporte. Existen tecnologías que utilizan el aprendizaje basado en similitudes: ante situaciones similares a las conocidas, se ofrecen soluciones similares.

Los fundadores creían que el desarrollo de la IA tendería a imitar a nuestro cerebro, pero la mayoría de las técnicas han evolucionado de forma muy distinta. En general, a día de hoy las tecnologías de IA requieren el procesamiento de cantidades ingentes de datos para dar respuestas.

Algunos de los **desafíos** que la IA tiene por delante son los siguientes: a día de hoy el análisis masivo de datos no ofrece explicabilidad. Los resultados se obtienen a través de cajas negras, sin que se pueda conocer el razonamiento que sustenta los resultados. Si queremos llegar a una IA fuerte, será necesario resolver esta cuestión. En segundo lugar, hasta la fecha no se ha conseguido integrar en un ordenador todos los componentes básicos de la inteligencia, es decir percepción, razonamiento, aprendizaje, comunicación y acción (aunque hay propuestas).

Es necesario integrar todos los componentes de la inteligencia. Finalmente, no se ha resuelto la desambiguación del lenguaje ni el correcto procesamiento de imágenes (percepción).

Una de las dificultades con las que se ha encontrado la IA es la imposibilidad de hacer una enumeración exhaustiva de todos los elementos que constituyen el conocimiento de sentido común con la idea de introducirlos en un ordenador. Una de las propuestas para dotar a un robot para que sea capaz de adquirir conocimientos por sí mismo es la denominada robótica del desarrollo ("developmental robotics"), que parte de la observación del sistema de aprendizaje de los niños y que se plasma en un modelo de computación del desarrollo mental para su implementación en un software.

**En resumen**, hasta la fecha no se ha resuelto el problema de la adquisición del conocimiento del sentido común (humanos) versus la inteligencia sistematizada (máquinas), y tampoco existe una base científica para creer que se va a lograr (frente a las tesis de la Universidad de la Singularidad).

Al finalizar la exposición del ponente se abrió un **debate**, en el que se abordaron las siguientes cuestiones:

- El cerebro necesita conocerse como fenómeno de autoprotección. Para ello puede utilizar la herramienta médica: construir tecnología para entenderse. Hoy, el cerebro humano se autodenomina consciente; aunque no sabe cómo ni porqué. Además, el cerebro es capaz de realizar abstracciones. Otra función emergente es lo que el mismo órgano denomina inteligencia, entendida como capacidad de resolver problemas inéditos, y que se sitúa en la corteza prefrontal. A continuación está lo que titula talentos (musical, para las matemáticas, etc.), que dependen del número de sinapsis y sobre todo de las interconexiones entre hemisferios. También se emociona y crea automatismos. Cada una de estas funciones se localiza en estructuras diferentes.
- Se plantea la cuestión de si la investigación en IA tiene en cuenta lo anterior, a lo que se responde señalando que existe una línea de investigación que se basa en el

funcionamiento del cerebro. Por ejemplo, los avances en fisiología de la visión (David Marr, MIT) han permitido el desarrollo de la visión artificial. Se ha conseguido reproducir los procesamientos eléctricos, pero lo que va más allá, como los procesamientos químicos, gliales y sinapsis, no se modelizan, no se tienen en cuenta.

- Se pregunta por el perceptrón y se habla del Memristor como el elemento básico para modelizar una neurona artificial. Su valor de resistencia depende del pasado. No es un elemento binario, sino que es capaz de variar su resistencia de forma o continua o analógica.
- Se plantea cuál es el fin último de la IA y cuál es el papel del ser humano. Siguiendo la exposición se diferencia entre los objetivos últimos de los fundadores, que aspiraban a construir inteligencias no humanas, guiados por la curiosidad científica y también para arrojar luz sobre cómo funciona el cerebro humano mediante la construcción de sistemas de IA, y los objetivos de la IA débil, que desde una visión más práctica persigue ayudar al ser humano a resolver problemas y tomar decisiones complejas. Se ha desarrollado una disciplina denominada "Ciencia cognitiva", que reúne a expertos en IA, en neurociencia, en filosofía, etc.
- En otra línea, se apunta que si partimos de la teoría del gen egoísta la IA no es más que una maniobra del cerebro humano para subsistir cuando el ser humano desaparezca.
- Existe una línea de investigación en IA enfocada a la modelización de emociones ("affective computing"). Hoy en día hay ordenadores capaces de identificar algunas emociones básicas mediante el análisis de imágenes de gestos faciales. Esto tiene aplicaciones en la enseñanza por ordenador, donde el ritmo o nivel de la clase se puede ir adaptando en función de la reacción del alumno.
- Respecto de la ironía o el humor, hasta la fecha no se ha conseguido que las máquinas lo perciban. Sobre el humor, hay dos investigadores en Trento (Italia) que han realizado varias publicaciones sobre la IA aplicada al humor (análisis de construcciones absurdas).
- Se pone de manifiesto que la aceleración de la investigación en IA es muy probable en una perspectiva de tres años, porque da dinero, por lo que la financiación y la inversión en IA se puede dar por asegurada. Frente a esta perspectiva, se argumenta que el auge actual de la IA ligado al no cumplimiento de las expectativas podría dar lugar en unos años al corte generalizado de la inversión, de manera similar a lo que ocurrió en los años 80. Lo que parece innegable es que los avances se dan de forma cada vez más rápida. Google es una de las empresas donde más investigación en IA se hace, lo que está llevando a la contratación de investigadores en las universidades, con el riesgo de vaciarlas. Esto es una preocupación creciente. Se considera positivo que la IA se dirija a aportar soluciones prácticas, pero se resalta que es importante no olvidar las grandes preguntas de la IA.

- Se apunta que la IA y su vertiente de tecnología cognitiva tienen una importante vinculación con lo social (y no sólo con lo comercial), en el sentido de que puede utilizarse para resolver problemas sociales, como los de las personas con limitaciones cognitivas.
- Se han realizado investigaciones sobre cómo dotar de sentido común a las máquinas, pero ha demostrado ser un campo extremadamente difícil. Las principales aproximaciones son el "crowdsourcing" (también llamado "wisdom of the crowds") y la robótica del desarrollo.
- Se cuestiona por qué la IA y la inteligencia humana deben ser sinónimos. Se responde a esto indicando que el cuerpo humano determina la forma en la que pensamos; así, si tuviéramos visión ultravioleta o si midiésemos 4 metros, pensaríamos de otra forma. Nuestro sistema de percepción determina nuestro cerebro. Por lo tanto, la IA será necesariamente distinta a la humana. En relación con la inteligencia humana, la investigación en neurología se está enfrentando actualmente al problema de la idea del libre albedrío.
- Se pregunta por los avances de la IA en el procesamiento del lenguaje natural, y si en la actualidad se ha superado el test de Turing (prueba de la habilidad de una máquina de exhibir un comportamiento inteligente similar, o indistinguible, del de un humano). Se comenta el caso del robot Eugene Goostman al que se sometió a dicho test simulando respuestas de un adolescente y del que se publicó que había superado el test de Turing. Sin embargo, se afirma que a día de hoy no se ha superado el test de Turing. Se cuestiona por qué Siri (Apple) funciona tan mal, aunque se precisa que funciona bien si se le dan instrucciones con frases sencillas. En relación con esto se informa de que el Gobierno en funciones acaba de adquirir un programa para el procesamiento del lenguaje natural que se desarrollará en idioma español.
- Se comparte la impresión de que, en general, se nos "vende" la IA débil como si fuera IA fuerte.
- Finalmente, se adelanta que en la próxima sesión sobre IA en el marco de los "Diálogos con la Ciencia", que tendrá lugar en junio, se abordarán cuestiones como el problema de la autonomía de la IA y la potencial sustitución del hombre por las máquinas.