Atlas filosófico de Inteligencia Artificial

Juan Miguel Aguado Terrón[1]

La aparición de la Inteligencia Artificial (IA) como proyecto experimental se encuentra perfectamente engarzada en la extensa tradición filosófica acerca de la cuestión central del conocimiento. Debido a ello, es posible abordar la historia y la sustancia de la IA conforme a las dos coordenadas que dibujan el espacio del pensamiento occidental: la abscisa delimitada por el par sujeto/objeto (el conocedor y lo conocido) y la ordenada delimitada por el par conocimiento/acción. La orografía característica en este territorio ha sido la fractura, la discontinuidad. A partir de la tradición eleática la idea de conocimiento empieza a tomar la forma de una representación. La metáfora lockeana de la mente como una habitación oscura en la que unas rendijas dejan pasar la luz del mundo exterior y con ellas las imágenes de las ideas y las cosas, marca una línea de pensamiento que va desde la caverna de Platón a las postrimerías de la IA, pasando por la Ilustración. Desde Kant, la cuestión central de la filosofía es saber cómo las representaciones de la mente se corresponden con las cosas del mundo. El lado mítico de esta trayectoria del pensamiento es la figura del autómata. Ésta nace como resultado de la primera fractura: una vez admitido el conocimiento como representación, y el sujeto como un representador universal, el paso al sueño de un sujeto artificial es meramente lógico. En este sentido, la codificación trascendental del sujeto llevada a cabo por Kant es crucial: Kant identificó y catalogó los componentes del sujeto cognoscente de modo tal que hacía posible soñar con reproducirlo. Mucho antes, en pleno siglo XIV, Ramón Llull había sentado sobre la lógica aristotélica las bases del sueño de un razonamiento perfecto y automático, el *Ars Magna*. La concreción lógica de tal sueño sería la empresa asumida por el positivismo lógico. La concreción experimental correspondería, más tarde, a las ciencias cognitivas y, en especial, a la IA. Esta trayectoria del autómata cognoscente puede ser denominada como el

[1] *Profesor de Fundamentos de la Información, Teoría de la Comunicación y Semiótica en la Universidad Católica San Antonio (Murcia). Fue profesor de Semiótica y Metodología de la Investigación en la Universidad de Wroclaw (Polonia) e investigador científico en el Instituto de Filosofía y Sociología de la Academia de Ciencias de Polonia (Varsovia).*

sueño de un sujeto sin sujeto, es decir, de un sujeto sin contexto, sin historia individual y social.

Paralelamente, la idea del conocimiento como representación introduce en la tradición occidental la metáfora de la visión como esquema dominante: *conocer es ver*. La premisa de un mundo objetivo intrínsecamente cognoscible e independiente de la actividad del sujeto (una colección de cosas percibidas, procesadas y catalogadas en la habitación oscura de la mente) suponía además una ulterior fractura: la establecida entre conocimiento y acción. En tanto que visión, el conocimiento era esencialmente pasivo, al menos en lo relativo a la percepción de las cosas en sí. El razonamiento (la actividad interna de la mente) quedaba entonces definido como un tipo especial de actividad distintivo del espíritu, un procedimiento automático de combinación de proposiciones, de acuerdo con la tradición llulliana. Esta disociación entre el conocimiento como acción interna del sujeto y la acción del sujeto en el mundo se encuentra en la raíz del pensamiento instrumental y sustenta, asimismo, una división genérica en la tradición de los autómatas. Es, pues, posible distinguir entre autómatas de la acción (el Golem, los criados de Hefestos, etc.) y autómatas del conocimiento (oráculos, autómatas jugadores de ajedrez, etc.). Unos y otros conforman corrientes separadas: los primeros desembocan en la moderna robótica (*robot* es la raíz común eslava para designar *trabajo*), los segundos en la IA.

El hecho de que el conocedor y lo conocido, tanto como conocer y hacer, hayan permanecido separados ha desembocado finalmente en la gran paradoja de las ciencias cognitivas: sujetos con historia se piensan en términos de sujetos sin historia a partir de la posibilidad de hacer sujetos sin historia. La crónica de las ciencias cognitivas y en especial de la Inteligencia Artificial es la crónica de la negación de sus orígenes: el reencuentro entre sujeto y objeto, entre hacer y conocer. El punto de partida de esta trayectoria es, como decía, el del pensamiento representacional.

El encuentro decimonónico entre la lógica y las matemáticas posibilitó la identidad operacional entre tres términos hasta entonces separados: representación, cálculo y computación. El álgebra de Boole y, especialmente, la Teoría Matemática de la Información de Shannon y Weaver, posibilitarían la hipótesis fundacional de la Inteligencia Artificial: pensar es computar, es decir, manipular representaciones simbólicas. Si existía un procedimiento universal del pensamiento (la computación), tenía que existir un operador universal de ese procedimiento (el sujeto artificial, el computador). Con la decisiva contribución del matemático George Boole se había consagrado a fines del XIX la identidad entre lógica y pensamiento (pensar equivale a articular proposiciones de modo formalmente adecuado). Con el aporte

shannoniano se consagraría la representación: la lógica booleana del todo o nada (unos y ceros) podía ser operada en circuitos eléctricos (conexión/desconexión) de modo que resultaba posible poner a prueba el carácter automático del pensamiento algorítmico o proposicional.

Un inconveniente quedaría pronto en evidencia: frente al mundo de la vida cotidiana, plagado de ambigüedades, mestizajes, dobles implicaciones y fronteras difusas, el mundo circunscrito por la IA se constituía sobre la rigidez formal de la bivalencia y las reglas estrictas de combinación. Si algo se puede definir (en el sentido matemático de una definición libre de ambigüedad), se puede representar, esto es, se puede computar. Tal era el manifiesto de la IA en sus primeros días. Desde el punto de vista experimental este camino no sólo conducía a, sino que partía de un callejón sin salida: la dificultad patente para afrontar la tarea de las definiciones (categorías, clases, distinciones temporales, formas, estructuras, etc.) ponía de manifiesto la distancia existente entre percepción y cognición.

Paralelamente a la corriente principal de la IA, la corriente denominada Reconocimiento de Patrones (RP) había centrado sus esfuerzos en el problema de la definición: ¿en qué consiste la esencia de un perro o de una casa? La cuestión kantiana de las categorías adquiría ahora rango experimental: si el mundo no era un tablero de ajedrez, ¿cómo establecer un procedimiento válido de definición y combinación de sus piezas? La vertiente instrumental apuntó a máquinas capaces de reconocer estructuras, tales como imágenes, colores, textos mecanografiados, etc. La vertiente reflexiva de la RP introducía en la IA el virus de un doble problema: percepción y aprendizaje. El resultado más conocido de este giro problemático en la IA son los *sistemas expertos*: sistemas que, a partir de un reducido número de datos y de un sistema de reglas de combinación, son capaces de integrar los resultados de su propia operación, esto es, en términos cotidianos, de *aprender*. En el período de gestación de la IA como opción experimental, durante la década de los 40, había surgido ya una llamada de atención sobre lo dificultoso de las relaciones entre el sistema y el medio (algo que, en términos cognitivos, afectaba de lleno a la doble fractura sujeto/objeto y conocimiento/acción).

El filósofo y matemático Warren McCulloch ideó las primeras *redes neuronales artificiales*, esto es sistemas de interconexión generalizada donde las operaciones lógicas eran *encarnadas* por conexiones de determinada clase. La hipótesis de trabajo implícita en la obra de McCulloch era doble: por un lado la presuposición de una cierta homología física y funcional entre los circuitos informacionales de Shannon (que más adelante harían posible el ordenador actual) y el funcionamiento de las neuronas conforme

a una lógica binaria; y, por otro, el de la emergencia de un orden lógico sobre el substrato de interconexiones generalizadas. La IA centraría su atención sobre la primera hipótesis, dejando de lado la segunda que, en cambio, constituye un foco de interés en la actualidad.

De modo simultáneo, basándose en los estudios realizados al filo de la Segunda Guerra Mundial sobre la autocorrección de los sistemas de disparo y guía de proyectiles así como en el modelo directriz de la homeostasis, Norbert Wiener, Arturo Rosenblueth y Julian Bigelow concibieron la cibernética como «la ciencia de la comunicación y el control en el animal y en la máquina». Aquella cibernética, ciertamente lejana de esa otra «cibernética» consagrada por los medios de comunicación masiva, retomaba el término platónico del *kybernetes*, el piloto o el timonel, para delimitar su objeto en torno a la *encarnación* del concepto de propósito (*purpose systems* o *problem solving machine* serían algunos de los nombres escogidos para bautizar a sus criaturas de laboratorio).

El matemático de origen húngaro John von Neumann, basándose en las aportaciones de McCulloch, Pitts, Wiener y los demás, sentó las bases para la concepción de los primeros *sistemas autónomos*: sistemas capaces de adaptarse y perdurar en permanente interacción perceptivo-activa con su entorno. En la concepción de estos sistemas resultaría fundamental la idea de emergencia, es decir, la *aparición* espontánea de un orden global a partir de coordinaciones locales generalizadas. Sin embargo, el auge que a finales de los años 60 adquirió el enfoque computacionalista clásico paralizó las investigaciones en este sentido, en parte debido al éxito de un diseño del propio Von Neumann que a la postre daría a luz al ordenador tal y como hoy lo conocemos. La idea del conocimiento quedaba, pues, momentáneamente circunscrita a los movimientos posibles en un tablero de ajedrez. Con todo, la contribución de aquella primera cibernética tendría un afluente imprevisto que acabaría por afectar al modo mismo de hacer ciencia. Con la teoría de sistemas esbozada por Ludwig von Bertalanffy a la luz de la cibernética wieneriana se abría un período de profundos cambios en el sentido que la ciencia otorgaba a la palabra «modelo». Si hasta entonces el modelo era un *simulacro de la realidad* en un sentido que ahora denominaríamos virtual, a partir de la IA la realidad se convertía en un simulacro del modelo. El modelo no era ya sólo lo que imitaba, sino lo que debía ser imitado: las computadoras pasaban de ser simulacros del cerebro y la cognición a ser modelos en el preciso sentido en que aquél (el cerebro) era entendido como un computador. Desde una perspectiva general, la idea de *modelización*, esto es la producción de modelos, recibía así carta de privilegio científico-tecnológico: la virtualidad entraba de lleno en la ciencia del