

JAVIER ELGUEA

Inteligencia artificial y psicología: la concepción contemporánea de la mente humana

It may be that the human mind is too complex to be understood by the human mind. But the desire to attempt the impossible seems to be one of its persistent features.

Aaron Sloman, 1978

Introducción

EN enero de 1982 el reconocimiento que la revista *Time* otorga anualmente al "hombre del año" fue concedido a una computadora. Esta es la primera vez que una máquina reemplaza a un ser humano en un galardón de esta naturaleza. La reacción que esta decisión provocó entre los lectores de la revista fue tan diversa como la actitud que los psicólogos manifiestan ante alguna mención a la Inteligencia Artificial: hubo airadas protestas, burlas, entusiastas cartas de aprobación, y expresiones de la más franca y abierta ignorancia.

El impresionante avance tecnológico que la humanidad ha experimentado en los últimos dos siglos encuentra su expresión más acabada y compleja en la invención de las computadoras, las "máquinas que piensan" como las llama McCorduck (1979). La rapidez y la asombrosa complejidad de la computadora ya no son un secreto para nadie, sin embargo la verdad sobre sus más profundas implicaciones es conocida sólo por un número muy limitado de especialistas.

Algunos expertos en computación han sostenido, desde hace algún tiempo, que las computadoras serán poseedoras de capacidades y habilidades similares a las de los seres humanos, y que, en el futuro próximo, las veremos igualarnos y superarnos en muchas de las actividades intelectuales tradicionalmente reservadas a los seres humanos. Con mucha mayor frecuencia que antes, expertos de varias disciplinas se preguntan: ¿puede una

máquina pensar?, ¿son las computadoras capaces de cognición?, ¿de experimentar emociones?, ¿existirá algo parecido a la mente humana en una computadora algún día?

El área que se ha dedicado durante las últimas tres décadas a estudiar y responder preguntas como éstas es conocida en círculos académicos como Inteligencia Artificial (IA). Hasta ahora, los logros y fracasos de la IA han sido causa de incontables controversias y enconados debates, de críticas acérrimas, y de grandes avances en disciplinas con intereses comunes como la psicología, la filosofía, la lingüística y la lógica.

El propósito de este artículo es el de exponer de una forma general y motivar a la comunidad académica de habla hispana, al diálogo entre psicólogos e investigadores en Inteligencia Artificial. En la redacción de este artículo, he intentado simplificar la discusión y mantenerla exenta de tecnicismos innecesarios. En la primera parte, se presenta una descripción de los objetivos de la IA, las presuposiciones teóricas y las herramientas metodológicas de esta nueva disciplina. En la segunda parte, se hace un breve recorrido por la historia de la IA, prestando especial atención a los eventos y trabajos que han tenido mayor importancia para la psicología. Finalmente, en la última parte, se discute la naturaleza y la importancia de la interacción entre IA y psicología y de sus contribuciones futuras.

¿Qué es la Inteligencia Artificial?

Como toda disciplina de reciente creación, la IA no se encuentra unificada en términos de objetivos y métodos de investigación. Recientemente, parte de los esfuerzos de los investigadores en esta área se han dedicado a la definición de dichos objetivos y al recuento de las herramientas metodológicas utilizadas hasta ahora (Boden, 1977; Dennett, 1978; Sloman, 1978; Ringle, 1979). Como resultado de este esfuerzo, que dista mucho de su conclusión, se han definido algunos acuerdos básicos sobre el área y sus estrategias.

Por ahora, es suficientemente claro que el objetivo de la IA es el de entender la naturaleza de la inteligencia a través del diseño de sistemas computacionales que la exhiban. En forma más concreta, puede afirmarse que, en lo que ha transcurrido de su corta historia, la IA ha estado dirigida por tres objetivos generales:

1. El análisis teórico de las posibles explicaciones del comportamiento inteligente
2. La explicación de habilidades mentales humanas
3. La construcción de artefactos (computadoras) inteligentes

Con estos propósitos en su agenda de investigación, los estudiosos de la IA han recurrido al uso de cuatro diferentes estrategias metodológicas: el desarrollo de tecnologías útiles en esta área, la simulación, el modelamiento, y la construcción de teoría sobre la inteligencia artificial. El desarrollo de tecnologías de computación ha sido una empresa titánica que los ingenieros en electrónica han tomado en sus manos, sin embargo, sólo una pequeña parte de lo que se conoce como ciencia de la computación puede incluirse dentro de la IA. No existe todavía un criterio preciso con el cual distinguir cuándo un sistema computacional es un sistema de IA, pero el acuerdo general es que cualquier máquina que desempeñe una función mental que tendría que ser realizada por una inteligencia humana es un ejemplo de IA.

La simulación que se hace en IA ha intentado reproducir algunas de las características inteligentes de los seres humanos. Estas reproducciones han buscado abiertamente la similitud entre una computadora y los seres humanos. La elaboración de simulaciones ha sugerido la posibilidad de explorar los procesos cognoscitivos humanos, sin embargo los esfuerzos en esta línea, a diferencia del modelamiento, han estado dedicados a producir comportamiento humano inteligente en las computadoras más que a entenderlo o explicarlo.

El modelamiento, por otra parte, tiene como objeto la utilización de los sistemas de IA para entender a la inteligencia humana. Ha sido tradicionalmente utilizado por psicólogos y no tiene como requisito necesario el uso de computadoras. De hecho, muchas de las teorías sobre cognición han utilizado modelos en computadoras sin hacer referencia a ellas, por ejemplo, la teoría sobre memoria semántica o sobre representación mental.

Finalmente, el trabajo teórico en IA ha abierto por primera vez la posibilidad de teorizar sobre la inteligencia sin hacer necesariamente referencia a la inteligencia humana. Es decir, se ha propuesto la formulación de una teoría de la inteligencia "pura".

Una breve incursión dentro de la historia de esta disciplina ilustrará más adecuadamente sus logros y su estado actual.

Breve historia de la Inteligencia Artificial

Los esfuerzos por reproducir algunas habilidades mentales humanas en máquinas y androides se remontan muy atrás en la historia. El mito del coloso de Rodas entre los griegos, las estatuas “parlantes” del medioevo, el androide de Von Kempelen que jugó al ajedrez con Napoleón, y el “motor analítico” de Charles Babbage que calculaba logaritmos, son sólo algunos de los ejemplos de este antiguo interés. Igualmente, la concepción de la inteligencia humana como un mecanismo no es reciente ni ha estado disociada de la psicología: Descartes, Hobbes, Leibniz, y el mismo Hume se refirieron a la mente humana como una forma de mecanismo.

Durante el siglo XIX y la primera mitad del XX, las analogías biológicas y fenomenológicas desplazaron a la noción de mecanismo en el estudio de la mente humana. Sin embargo, a partir de la segunda mitad de nuestro siglo, la noción de mecanismo renovó su poder heurístico con la formalización de la noción de “computación”.

Como algunas máquinas, especialmente las calculadoras, se diseñaron para evitar el tener que pensar y para hacer el pensamiento más rápido y exacto, fue inevitable que desde sus orígenes las calculadoras, y más adelante las computadoras, se relacionaran con la inteligencia y el pensamiento enfatizando sus similitudes.

La IA fue introducida a la comunidad científica en 1950 por el inglés Alan Turing en su artículo “Maquinaria Computacional e Inteligencia.” A pesar de que la investigación sobre el diseño y las capacidades de las computadoras comenzaron algún tiempo antes, fue hasta que apareció el artículo de Turing que la idea de una máquina inteligente cautivó la atención de los científicos.

La pregunta básica que Turing trató de responder afirmativamente en su artículo era: ¿pueden las máquinas pensar? Los argumentos de Turing en favor de la posibilidad de inteligencia en las máquinas, iniciaron un intenso debate que marcó claramente la primera etapa de interacción entre la IA y la psicología. Los debates en aquella época se centraron en el análisis de la serie de problemas implicados en la aplicación de terminos mentalistas a las computadoras. La intención de Turing no era la de usar estos términos como analogías sino la de eliminar la distinción entre inteligencia natural e inteligencia artificial.

Dos de las contribuciones más importantes de Turing a la IA fueron el diseño de la primera computadora capaz de jugar ajedrez y, más importante que esto, el establecimiento de la naturaleza simbólica de la computación.

El trabajo de Turing, quien falleció prematuramente, fue continuado en los Estados Unidos por John Von Neumann durante la década de los cincuentas. Su contribución central fue la idea de que las computadoras deberían diseñarse tomando como modelo al cerebro humano. Von Neumann fue el primero en “antropomorfizar” el lenguaje y la concepción de la computación al hablar de la “memoria”, los “sensores”, etc., de las computadoras. Construyó una serie de máquinas utilizando lo que a principios de los cincuentas se conocía sobre el cerebro humano, y diseñó los primeros programas almacenados en la memoria de una computadora.

Sin embargo, esta línea de investigación pronto encontró serias limitaciones. La concentración en la imitación de la constitución físico-química del cerebro, no permitió ver, a Von Neumann y sus seguidores, que la analogía sería mucho más eficiente si se estudiaran las funciones del cerebro, es decir, sus capacidades como procesador de información.

Corresponde a McCulloch, a mediados de los cincuentas, formular una posición radicalmente distinta al sostener que las leyes que gobiernan al pensamiento deben buscarse entre las reglas que gobiernan a la información y no entre las que gobiernan a la materia. Esta idea abrió grandes posibilidades a la IA. En esta línea, Minsky (1959), uno de los padres fundadores de la IA, modificó su posición y sostuvo que la imitación del cerebro a nivel celular debería ser abandonada.

Es más o menos en esta época que ocurre un evento que organizaría y daría un gran impulso al desarrollo de la IA: el congreso en Dartmouth (1956). En este congreso, en el que se reunieron los padres fundadores de la disciplina, se llegó a la definición de las presuposiciones básicas del núcleo teórico de la IA:

1. El reconocimiento de que el pensamiento puede ocurrir fuera del cerebro, es decir, en máquinas
2. La presuposición de que el pensamiento puede ser comprendido de manera formal y científica
3. La presuposición de que la mejor forma de entenderlo es a través de computadoras digitales

Desde fines de los cincuentas la investigación en IA se expande y se multiplica en direcciones diversas. La capacidad simbólica de las computadoras es estudiada, entre otros, por Shanon (1950) y por Newell, Shaw y Simon (1958) quienes diseñan el primer programa inteligente basado en su modelo de procesamiento de información. Este modelo de Newell, Shaw y Simon habría de convertirse pronto en la teoría dominante en psicología cognoscitiva.

Algunos investigadores se dedicaron al estudio de la naturaleza del aprendizaje en las computadoras y a los procesos de reconocimiento de patrones visuales. Como resultado de ello Selfridge y Dinneen consiguen diseñar el primer programa capaz de aprender por experiencia (ver McCorduck, 1979).

Basándose en los estudios sobre memoria asociativa, el equipo Newell-Shaw-Simon construyó los primeros lenguajes de procesamiento de información (IPL-I, IPL-II) utilizados en el diseño de su "Logic Theorist Machine" que se convirtió en la primera máquina "inteligente". Esta máquina fue capaz no sólo de memorizar y aprender, sino que consiguió demostrar de una manera original y "creativa", es decir no prevista por sus creadores, algunos de los teoremas propuestos por Bertrand Russell en los Principios (Russell and Whitehead, 1925).

Desde sus orígenes la IA se relacionó con juegos como el ajedrez y las damas, probablemente debido a que los juegos de mesa constituyen modelos de situaciones reales en las que hay que calcular, solucionar problemas, tomar decisiones, corregir errores, recordar, etc. A pesar de que esta línea de investigación ha sido casi totalmente abandonada en la actualidad, muchos de los avances teóricos y metodológicos de la IA se deben a ella. Por ejemplo, Samuel diseñó en 1961 un programa que jugaba damas y que era capaz de aprender de sus errores, es decir, era capaz de adaptar su comportamiento en relación a eventos pasados. Lo pasmoso de este programa fue que, aunada a su capacidad de aprendizaje la de memoria, con el tiempo consiguió derrotar invariablemente a su creador. El mismo resultado fue obtenido por Bernstein a través de un programa que jugaba ajedrez (Boden, 1977). Los grandes "retos" entre computadoras y seres humanos se multiplicaron, siendo el más famoso de ellos el que ocurrió entre Dreyfus (un enconado crítico de la IA) y el programa Machack, en el que Dreyfus fue derrotado en un juego de ajedrez de varias horas.

A principios de los sesentas, la IA comienza una fase distinta de su desarrollo. En 1962, McCarthy y Raphael inician sus trabajos sobre el diseño y la construcción de un robot móvil que llamarían "Shakey". La diferencia fundamental entre este robot y los programas en computadora utilizados hasta ahora por la IA, es que "Shakey" tendría que enfrentar el reto de interactuar con el mundo real en términos de espacio, tiempo, movimiento, etc. En otras palabras, el robot tendría que tener alguna forma de "conocimiento" del mundo que lo rodeaba. Este reto inició una fuerte preocupación en la IA por el estudio de la epistemología y los procesos cognoscitivos. La discusión se centró alrededor de los problemas de la representación mental o interna del conocimiento, la percepción y los problemas del significado. La idea básica de Raphael era la de reunir, en una sola, distintas máquinas con capacidad de aprender por experiencia, de reconocer patrones visuales, de modelar, de manipular símbolos, etc., y esperar que el todo fuera mayor que la suma de las partes. El resultado del experimento no fue el éxito que Raphael esperaba, pero fue un logro sin precedente que hizo posibles avances muy importantes. El aprendizaje más importante de esta experiencia fue la comprensión de que el problema más difícil por resolver en IA era el de construir una máquina capaz de funcionar con altos niveles de incertidumbre, como lo hace un ser humano. Se hizo claro que construir una máquina que no lidiara efectivamente con la incertidumbre sería una de dos: o trivial, por la simpleza de la tarea, o imposible por la complejidad de la misma.

Hacia mediados de los sesentas la IA se convierte en un área en la que se interesan e interactúan especialistas de diversas disciplinas: lógicos, psicólogos, matemáticos, lingüistas, filósofos, etc. Uno de los grandes temas de IA en esta década fue el estudio del lenguaje. En la mayoría de los estudios iniciales sobre lenguaje, se atacó el problema de diseñar una máquina que fuera capaz de traducir de un idioma a otro. El énfasis se hizo en el análisis de la sintaxis, en lugar del significado, estrategia que se abandonó relativamente pronto. Los investigadores interesados en esta área de la IA pronto descubrieron que traducción no es lo mismo que transformación, y que, como consecuencia de ello, de alguna manera la máquina tendría que "entender" un texto antes de poder traducirlo. Los esfuerzos se orientaron hacia una definición de la comprensión que pudiera simularse en una computadora.

Con esta idea como lineamiento heurístico, se desarrolló una serie de programas que exploraban esta serie de habilidades relacionadas con el lenguaje y la comprensión: el "oráculo" de Phillips, "Sad Sam" de Lindzay que fue uno de los más exitosos, y "Eliza" de Wizenbaum y Colby. (ver Boden, 1977). El programa de Lindzay se orientó hacia el estudio del significado, connotativo y denotativo, y es capaz de hacer inferencias, conocer, y parafrasear sobre el mundo conocido. Eliza, por otra parte, consistió en un análisis de la conversación que se basa en la simulación de una terapia psicológica interactiva en la que el programa es el analista y el interlocutor es el analizado. Un ejemplo de una conversación sostenida con Eliza en la computadora de la Universidad de Stanford es el siguiente (el diálogo de la computadora en letras mayúsculas):

BUENAS TARDES, POR FAVOR ESCRIBA SU NOMBRE
 buenas tardes, me llamo Javier
 ES UN PLACER CONOCERTE, JAVIER
 el placer es mio
 ¿QUE TE TRAE POR AQUI?
 nada en especial, he tenido trabajo y me siento agotado
 ¿QUE TE HACE PENSAR QUE ESTAS AGOTADO?
 he tenido problemas en mi trabajo
 ¿PROBLEMAS?
 es decir, he estado preocupado con mi trabajo
 CUENTAME TUS PREOCUPACIONES...

Según Colby, psicoanalista por entrenamiento, el programa resultó tan exitoso en términos clínicos que valdría la pena intentar reproducirlo masivamente como instrumento terapéutico. Intento al que Weizenbaum, coautor del programa, se opuso decididamente con el argumento de que no puede decirse desde ningún punto de vista que "Eliza" es capaz de entender lo que se le dice. "Eliza", Weizenbaum insiste, es un programa capaz de responder, que no es lo mismo que comprender.

Como resultado de este y otros debates, las preguntas de investigación siguieron girando alrededor de la noción de comprensión. Los seres humanos comprendemos en una plática mucho más de lo que se nos dice explícita y verbalmente (en parte porque la comunicación no-verbal juega un papel muy importante y en parte por nuestras capacidades para obtener información del contexto y para operar con altos niveles de

incertidumbre) y pareció importante que estas habilidades fueran reproducidas en una computadora con pretensiones de inteligencia. Los investigadores se dedicaron entonces al estudio de los aspectos no-verbales del lenguaje y al de reconocimiento de patrones.

Para mediados de los sesentas, y como resultado de los estudios sobre lenguaje y comprensión, la IA se involucró en el estudio del problema de la representación mental del conocimiento. Igualmente, una gran preocupación en esta década fue la de encontrar aplicaciones prácticas de lo que se había aprendido hasta entonces. Entre las aplicaciones más exitosas debe mencionarse el diseño de DENDRAL en Stanford, un programa que hace espectogramas y diagnósticos médicos y que hasta ahora ha tenido la misma tasa de error que los seres humanos. La segunda aplicación importante tiene que ver con la psicología educativa y la mencionaré en la siguiente sección sobre la interacción entre la IA y la psicología.

La interacción entre IA y psicología: teoría y método

A pesar de que la interacción entre IA y psicología no es ya ninguna novedad, los mejores frutos y resultados de esta interacción están aún por venir.

Haciendo a un lado la importante preocupación que los especialistas en IA tienen por la construcción de sistemas electrónicos de computación, es evidente que la psicología y la IA se empalman en gran medida. Es decir, ambas disciplinas están preocupadas con la explicación y resolución de los mismos problemas básicos, sus soluciones pueden ser comparadas una contra otra, y su éxito o fracaso puede evaluarse con el mismo criterio. El objetivo de la IA y la psicología es el de entender aspectos diversos de la mente humana y de la inteligencia en general y, en esencia, el juicio final que tendrán que enfrentar tendrá que ver con la cantidad y la calidad de las explicaciones que proporcionen sobre la naturaleza de la actividad mental.

Esto no debe hacernos verlas, como algunas personas han sugerido, como rivales teóricas necesariamente. Por el contrario, yo sostengo que, hasta ahora, la psicología ha sido la mejor y más rica fuente de teorías y conjeturas para la IA y, a su vez, que la IA es la mejor medicina que la psicología necesita en el estadio actual de su desarrollo.

Trataré de ilustrar esta última afirmación en esta sección con argumentos teóricos, metodológicos, y prácticos.

En primer lugar los argumentos de carácter teórico. Una de las características de las teorías científicas progresivas (Lakatos, 1978) es el equilibrio que mantienen entre la generación de conjeturas o hipótesis teóricas y su corroboración o refutación empírica. Una gran parte del desarrollo de la psicología se ha caracterizado por la ausencia de este equilibrio debido, entre otras causas, a que la experimentación con seres humanos se encuentra limitada por importantes razones morales. El estudio de la actividad mental ha tenido que limitarse a los reportes introspectivos, a pesar de que es sabido que la mayor parte de la actividad mental no es consciente y está fuera del alcance de la introspección. Las consecuencias de esto son, por un lado, que existe un gran divorcio entre teoría y realidad empírica como lo ejemplifica la psicología clínica, y por el otro, que la teoría en psicología experimental se ha subdividido en una innumerable cantidad de micro-modelos psicológicos de tan limitado alcance y generalizabilidad que en muchos casos su aplicación se restringe únicamente a las condiciones experimentales (Cronbach, 1975). Finalmente, el efecto general es que la teoría no es dirigida por lineamientos heurísticos como en las ciencias maduras sino por una estrategia burda de ensayo y error.

Es mi opinión que, en su intento por reproducir habilidades mentales humanas, los investigadores en IA han creado sistemas teóricos que poseen una gran capacidad heurística, que reúnen todos los requisitos científicos de consistencia y rigor lógico, y que permiten, si no responder de manera conclusiva, sí al menos formular mejores preguntas psicológicas.

Algo que hemos aprendido a lo largo de la historia de la psicología es que sus objetivos iniciales son más difíciles de alcanzar de lo que originalmente se pensaba y que se requieren para ellos mejores herramientas metodológicas y mejores teorías.

Hoy en día, uno de los más importantes objetivos de la psicología es el de explicar fenómenos como la percepción, el aprendizaje, la cognición, la motivación, etc., de forma tal que sea posible unir las teorías fisiológicas a las psicológicas y terminar con el dualismo que ha plagado estas áreas de estudio.

Para ello se ha recurrido a dos estrategias experimentales distintas: La primera de ellas, que llamare "inductivista", comienza al nivel neuronal o atómico, unifica estas unidades físicas elementales en forma de moléculas, que a su vez reúne en

forma de grandes agregados, que eventualmente describen el comportamiento y los fenómenos psicológicos que todos podemos observar. Esta estrategia, dicho sea de paso, ha avanzado muy lentamente y tenido un éxito sumamente limitado en la explicación de los fenómenos que van más allá de la mera percepción sensorial. La segunda estrategia, que llamaré “deductivista”, comienza con una descripción de las funciones y organización de los procesos psicológicos, y descompone estos sistemas en sistemas más simples y pequeños, con la esperanza de eventualmente llegar a describir elementos fisiológicos y físicos.

Es importante aclarar que ambas estrategias deben ser continuadas ya que ambas han contribuido hasta ahora con parte de las explicaciones existentes, sin embargo, la evidencia sugiere que la primera estrategia es de poca utilidad para la psicología. La ilustración de esto es muy simple en términos de preguntas empíricas. Es mucho más difícil responder a la pregunta: ¿cómo puede un número determinado de neuronas (o átomos) producir fenómenos como inteligencia, cognición, lenguaje, toma de decisiones, etc.?, que responder a la pregunta ¿cómo puede un sistema (ser humano o computadora digital) con determinadas características realizar tal o cual tarea o función?

La segunda pregunta busca una solución, cualquier solución, la primera requiere de un descubrimiento. La búsqueda de descubrimientos por mera inducción o “inspección bruta” puede ser interminable (si consideramos, por ejemplo, el número de combinaciones neuronales posibles) si uno no toma como dirección heurística alguna pregunta psicológica (algo que, a pesar de todo, los neurofisiólogos hacen tácitamente y sin gran competencia). Y, en última instancia, la búsqueda de soluciones puede en ocasiones llevarnos a un descubrimiento, es decir, a encontrar la solución que la naturaleza encontró primero.

Las teorías en IA caen dentro de la categoría deductivista y, por lo tanto, poseen una estrategia con mayor capacidad explicativa. Debo insistir, ambas estrategias son necesarias pero la experiencia indica que la estrategia deductivista ha proporcionado y continuará proporcionando la mejor y mayor parte de las explicaciones. Esto, dicho sea de paso, es cierto también en otras ciencias, la estrategia inductivista ha sido abandonada en ciencias como la química y la biología que han favorecido un enfoque deductivista. De tal forma que, si el propósito es seguir el ejemplo de las ciencias “maduras”, el camino es claro.

Las contribuciones que esta estrategia, típica de la teoría en IA, han traído a la psicología son varias, mencionaré sólo algunas de las más importantes.

Una de las contribuciones centrales de la IA se relaciona con un problema que ha agobiado a psicólogos y filósofos por más de doscientos años: el problema de la representación interna. Por mucho tiempo ha sido evidente que la única psicología capaz de dar cuenta de las complejidades de la actividad mental tendría que explicar e incluir procesos de representación interna. Por ejemplo, es la opinión de la mayoría de los psicológicos que sería imposible explicar el fenómeno de la percepción si no presuponemos que su función es proveernos con una imagen mental o modelo del mundo en el que vivimos y actuamos de una forma más o menos exitosa.

Esta premisa de la representación interna ha sido aceptada por todos los psicólogos con la excepción de los conductistas radicales como Watson o Skinner. Las representaciones internas han sido denominadas hasta ahora de formas distintas: ideas, sensaciones, impresiones. Más recientemente nos referimos a ellas como esquemas cognoscitivos, mapas mentales, proposiciones, imágenes, o señales neuronales. Sólo este breve recuento de las distintas denominaciones de las representaciones mentales hace obvias dos conclusiones: primero, el acuerdo existente sobre la necesidad de incluirlas en nuestras explicaciones de la actividad mental, y, segundo, el desacuerdo general sobre su naturaleza y características.

El serio problema que la representación mental ha impuesto a la teoría psicológica es el de explicar de que forma estas representaciones pueden modelar, por sí mismas, al mundo en la mente humana. En otras palabras, es claro que nada puede ser intrínsecamente una representación de otra cosa. Algo se convierte en una representación sólo *para* alguien, todo sistema de representación requiere de un intérprete, independiente de este sistema, que lo utiliza con algún propósito. Este intérprete debe poseer una gran variedad de atributos psicológicos: debe poseer intenciones, debe ser capaz de comprender, debe tener objetivos y creencias, de tal forma que sea capaz de utilizar las representaciones para informarse sobre el mundo y ayudarse en el logro de sus metas (conversar, resolver problemas, reconocer, etc.). Este intérprete dentro de nuestra mente, es, en esta concepción, una forma de "homúnculo", y los homúnculos como lo sabemos desde Descartes, padecen el mal de la regresión infinita. Lo que

quiere decir que, al afirmar la existencia de un homúnculo dentro de la mente humana, caemos en la necesidad de recurrir a otro homúnculo dentro del primero que cumpla la misma función. Del mismo modo, más adelante necesitaremos de un tercero dentro del segundo... así *ad infinitum*. Esta salida *ad hoc*, porque evidentemente no es una solución, no ha hecho más que transferir el problema de la representación de la mente humana a otra distinta, de tal forma que, como sostiene Dennett:

Hacer psicología sin homúnculos ha sido imposible. Pero hacerla con homúnculos nos ha llevado a la circularidad y el regreso infinito, al punto que ha convertido a la psicología en una empresa imposible. (Dennett, 1979:71)

El reto ha sido entonces, formular una teoría que explique cómo pueden las representaciones entenderse a sí mismas y, al mismo tiempo, evitar la circularidad y el regreso infinito.

Como lo mencioné anteriormente, por dos siglos se ha tratado de resolver este problema sin ningún éxito. No es sino hasta la aparición de la noción de "estructuras de datos" dentro de la IA que el problema parece comenzar a resolverse. Las "estructuras de datos" pueden o no tener validez psicológica, es algo que está aún por aclararse, lo que ya es claro es que constituyen el primer y único ejemplo operante de representaciones que se "entiendan a sí mismas". Las estructuras de datos han permitido la construcción de sistemas intencionales que son resultado del diseño de subsistemas con habilidades específicas y simples que se encuentran organizados entre sí. Estos subsistemas se subdividen a su vez en microsistemas más especializados y simples, que a su vez se componen de sistemas más pequeños y tan simples que sus respuestas son binarias. El sistema total consigue modelar una forma operante de representación.

El programa diseñado por Winograd (1972), conocido como SHRDLU es un extraordinario ejemplo de este comportamiento. SHRDLU es un programa que vive en un mundo imaginario de figuras geométricas en color, que es capaz de manipular estas figuras sobre una superficie cuando se le solicita, que representa a sí mismo este mundo, y que puede responder preguntas sobre su vida en este mundo. Un ejemplo de un diálogo con SHRDLU es el siguiente (de nuevo las respuestas de la computadora en mayúsculas):

¿Tocaste alguna pirámide antes de tocar el cubo azul?

SI, LA VERDE

¿En qué momento?

MIENTRAS APILABA EL CUBO ROJO PEQUEÑO ENCI-
MA DEL CUBO ROJO GRANDE Y DEL CUBO VERDE

¿Por qué la tocaste?

PORQUE NECESITABA MOVERLA PARA COLOCAR EL
CUBO ROJO PEQUEÑO EN SU LUGAR

¿Para qué?

PARA PONER EL CUBO ROJO PEQUEÑO DONDE ESTA
AHORA

¿Por qué?

PORQUE TU ME LO PEDISTE...

Las respuestas del programa son una muestra de la claridad con la que representa y opera en su mundo imaginario, incluyendo al interlocutor que da las órdenes, basándose únicamente en esta idea de estructuras de datos. Ninguna otra explicación ha llegado tan cerca de modelar el proceso de representación interna.

Otro problema en el que la IA ha sido de gran utilidad teórica se refiere a la discusión sobre la existencia y la naturaleza de las imágenes mentales. He discutido este problema en detalle en otro artículo (Elguea, 1983) de manera que referiré al lector interesado a este artículo y sólo mencionaré que el problema de las imágenes mentales se encuentra íntimamente relacionado con el de la representación y que el trabajo en IA ha inclinado a la teoría psicológica en favor de la eliminación de las imágenes como concepto teórico y la aceptación de las proposiciones como el vehículo del pensamiento y la representación.

En segundo lugar, analizaré las contribuciones metodológicas de la IA a la psicología.

Algo que no ha sido comprendido adecuadamente entre los psicólogos, a veces ni siquiera entre especialistas en AI, es que los programas de IA no son experimentos empíricos, sino experimentos "pensados" o imaginarios regulados rigurosamente por una computadora.

La lógica de programación por computadora ha desarrollado una serie de reglas de consistencia que proveen al psicólogo de un lenguaje técnico con qué disciplinar la formulación de teorías y conjeturas, además de proporcionarnos nuevas y mejores metáforas para estudiar la inteligencia.

Los programas que operan en una computadora y la programación en general nos proveen de un nuevo medio para generar ideas, pensar acerca de ellas, probarlas, extenderlas, corregirlas, profundizarlas y, eventualmente, explicarlas. Adecuadamente programadas, las computadoras son instrumentos útiles para construir, manipular, analizar, interpretar y transformar estructuras simbólicas de cualquier tipo, incluyendo sus propios programas.

Igualmente, persiste entre la comunidad científica el mito de que el estudio de sistemas complejos requiere necesariamente del uso de mediciones numéricas, ecuaciones, cálculo, estadística, etc., que las ciencias sociales han heredado de la física. Estos instrumentos han demostrado su insuficiencia en la explicación de fenómenos esenciales a la actividad mental de cualquier tipo, inclusive de la actividad "mental" que puede ocurrir en una computadora. (Ningun especialista en computación intentaría entender el comportamiento conversacional o cognitivo de un programa en términos estadísticos o de variables dependientes e independientes).

En lugar de ecuaciones, sistemas también formales no-numéricos han sido desarrollados en la forma de lenguajes de programación. Los conceptos generados a través de este esfuerzo, como lo ha demostrado el modelo de procesamiento de información, pueden resultar de mucha mayor utilidad para el psicólogo que las matemáticas numéricas que se utilizan normalmente y que tienen una capacidad limitada para teorizar sobre estructuras complejas e interactivas como las que se dan en la actividad mental. Es desafortunado que, en la mayoría de los casos, los psicólogos tengan un escaso o nulo entrenamiento en computación o programación y que su información se limite al uso de programas estadísticos escritos en lenguajes simples como FORTRAN o BASIC, y que ignoren la existencia de lenguajes más poderosos para la simulación y el modelamiento de actividades mentales como son los casos de LISP, y los más recientes CONNIVER, PLANNER, KRL, etc.

Por último, en lo referente a las aplicaciones prácticas, la IA ha contribuido a la psicología a través de la creación de programas dedicados a la educación de niños. Como uno de los resultados del esfuerzo por entender, simular, y modelar habilidades mentales, Papert (1971) desarrolló una serie de lenguajes conocidos como LOGO que tienen por objeto enseñar estas mismas habilidades a los niños. Estos lenguajes tienen grandes

ventajas sobre otros, en primer lugar son mucho más poderosos y por lo tanto hacen mucho más fácil la tarea de programar a la computadora para realizar actividades complejas, y, en segundo lugar, no se restringen al uso de computaciones numéricas. Por ejemplo, LOGO y POP-2 (utilizados en MIT y en Sussex respectivamente) son lenguajes que tienen la capacidad de manipular palabras y oraciones, figuras geométricas, diseño de imágenes, etc.

La oposición en contra de la IA no ha sido poca ni sin importancia. Dreyfus (1972), el más importante crítico de la inteligencia artificial, ha elaborado serias críticas basándose en la distinción entre computación análoga y computación digital. Dado que todas las simulaciones y modelos de la cognición son generados en computadoras digitales, los procesos involucrados utilizan estados de información discreta y programación secuencial. Dreyfus insiste en que la evidencia que ha provisto la psicología y la neurofisiología, indica que el cerebro humano funciona como una computadora análoga que utiliza operaciones continuas y estructuras de programación en paralelo. El sostiene que, como resultado de ello, las simulaciones y modelos generados en la IA carecen de validez psicológica.

Si bien es cierto que la crítica de Dreyfus es aplicable a un buen número de modelos y simulaciones, pierde de vista la capacidad que tiene una computadora digital para simular el funcionamiento de una computadora análoga (por ej. el cerebro), y que su argumento lejos de refutar los esfuerzos hechos hasta ahora por la IA los ha fortalecido indirectamente.

En la actualidad muchas de las nociones y términos utilizados en la psicología tienen su origen en modelos mecanicistas o, físicos, insuficientes y simplistas. Inclusive el grupo de académicos que se oponen fuertemente a la utilización de metáforas computacionales para describir procesos mentales, con el pretexto de que son nociones mecanicistas, usan analogías mucho más crudas como por ejemplo: "disonancia", "represión", "tensión", "carga emocional", "fuerza motivacional", etc. Los oponentes de las metáforas computacionales de alguna manera están condenando a la teoría psicológica a continuar viviendo con analogías hidráulicas y mecánicas derivadas de avances tecnológicos y científicos de algunos siglos atrás.

Nota final

Existe todavía una enorme cantidad de fenómenos psicológicos sin descripciones ni explicaciones adecuadas. Entre ellos: percepción, memoria, reconocimiento de patrones, comprensión, aprendizaje, comunicación, interpretación, creación, intuición, selección, etc. Los psicólogos, como la mayoría de los científicos sociales, han tenido hasta ahora un conjunto de herramientas inadecuadas para la formulación de teoría y la corroboración de hipótesis sobre esta serie de fenómenos. Algunos de ellos han tratado de recurrir sin gran éxito a modelos más recientes como la teoría de sistemas o la teoría de juegos que, por su naturaleza estática, siguen siendo deficientes en la explicación de la actividad mental.

La IA ha explorado las distintas formas en que las computadoras podrían realizar las tareas que antes estaban reservadas a los seres humanos, como resolver problemas, planear a futuro, demostrar teoremas, jugar ajedrez, conversar en y entender un lenguaje, componer música, etc. El hecho de que aún no haya conseguido reproducir un ser humano completo (o que eventualmente lo consiga) es de menor importancia que la evidencia de que ha mejorado nuestras habilidades para pensar y clarificar fenómenos de interés para la psicología y otras ramas de la ciencia. Ha colaborado en la reformulación de viejos problemas psicológicos y en la reconsideración y reevaluación de las teorías existentes. Ha forzado a los especialistas a precisar sus conceptos y a utilizarlos con mayor consistencia y rigor, al mismo tiempo, los ha enfrentado a un reto teórico al formular un programa rival de investigación que busca explicar un objeto de estudio que anteriormente estaba reservado a los psicólogos: la explicación de la inteligencia y sus procesos colaterales.

Sus logros y avances han sido tan importantes que, en unos cuantos años, las teorías sobre la inteligencia y sus fenómenos relacionados serán literalmente incomprensibles para el psicólogo que no esté familiarizado con el desarrollo teórico, los métodos y los logros de la IA. Su influencia en la psicología ha sido y será tan grande que enseñar cursos universitarios sobre psicología cognoscitiva sin hacer referencia a la IA será un grave acto de ignorancia o de irresponsabilidad.

Ahora bien, con respecto al objetivo último, no alcanzado, de la IA, el de entender y reproducir al ser humano, vale la pena mencionar de nuevo las palabras del profesor Sloman:

Puede ser que la mente humana sea demasiado compleja para ser comprendida por la misma mente humana. Pero el deseo de intentar lo imposible parece ser una de sus características más persistentes.

(Sloman, 1978:20)

BIBLIOGRAFÍA

- BODEN, M. *Artificial Intelligence and Natural Man*. New York: Basic Books, 1977.
- CRONBACH, J.L. "Beyond the Two Disciplines of Scientific Psychology", Febrero 1975, en *American Psychologist*, Vol. 30, No. 2, 116-127.
- DENNETT, D.C., *Brainstorms*. Vermont: Bradford Books Publishers Inc, 1978.
- DREYFUS, H., *What Computers Can't do: A Critique of Artificial Reason*. New York: Harper, 1972.
- ELGUEA, J.A., "El Debate Sobre Representación Mental en Psicología Cognoscitiva: una Reconstrucción Racional", 1983, por aparecer en *Journal or Interamerican Psychology*.
- LAKATOS, I., *Philosophical Papers*, London: Cambridge University Press, 1978.
- McCORDUCK, P., *Machines Who Think*. San Francisco: Freeman and Co, 1979.
- McCULLOCH, W., *Embodiments of Mind*. Cambridge, Mass.: MIT Press, 1965.
- NEWELL, A., SASHW, J.C., AND SIMON, H.A., "Chess Playing Programs and the Problem of Complexity", 1958, en *IBM Journal of Research and Development*, 2, 4.
- and SIMON, H.A., "The Logic Theory Machine", septiembre 1956, en *IRA Transactions in Information Theory*.
- PAPERT, S., *Teaching Children Thinking*. AI memo 247, MIT: AI Laboratory, 1971.
- PYLYSHYN, Z. W., "Complexity and the Study of Human and Artificial Intelligence", en *Philosophical Perspectives in Artificial Intelligence*, Ringle, M.D. (ed), England: Humanities Press, 1979.
- RINGLE, M.D., *Philosophical Perspectives in Artificial Intelligence*. England: Humanities Press, 1979.
- RUSSELL, B. AND WHITEHEAD, A.N., *Principia Mathematica*. England, Cambridge University Press, 1925-1927.
- SHANON, C., "A Chess Playing Machine", February 1950, en *Scientific American*, 182.
- SLOMAN, A., *The Computer Revolution in Philosophy*. England: Humanities Press, 1978.
- TURING, A., "Computing Machinery and Intelligence" (1950), in *Computers and Thought*. Feigenbaum, E.A. and Feldman, J. (eds.), New York: McGraw Hill, 1963.
- WINOGRAD, T., *Understanding Natural Language*. New York: Academic Press, 1972.