LA INTELIGENCIA ARTIFICIAL A TRAVÉS DE SUS CIENTÍFICOS

Alberto Suárez González
Dpto. Ingeniería Informática. Universidad Autónoma de Madrid

RESUMEN

Las raíces de la inteligencia artificial son más antiguas y profundas de lo que convencionalmente se presupone en el relato canónico. Avanzaremos la hipótesis de que la creación de artefactos que muestran comportamientos inteligentes es la continuación natural de la búsqueda del *logos*, el principio objetivo que gobierna el cosmos, que permite al humano conocer y actuar sobre el mundo. Siguiendo la tradición filosófica, lo haremos hablando de los científicos que han hecho posible esta apasionante empresa.

1. DE CÓMO EL HUMANO IDEÓ IMPLANTAR INTELIGENCIA EN LA MÁQUINA Y DE LAS DIFICULTADES QUE ENCONTRÓ EN SU EMPEÑO.

El anhelo del humano por concebir en su mente, diseñar gracias a su ingenio, y construir con sus manos un sistema artificial que emule, cual ilustrado reflejo de la suya, la elusiva cualidad de la inteligencia es probablemente simultáneo a su deseo de comprender el mundo, al otro y a sí mismo. A lo largo de la historia el proyecto adopta disfraces varios: la reconstrucción del olvidado hebreo mediante el cual Dios se comunicó con Eva y Adán: un lenguaje tan certero que, por su propia perfección, no diera cabida al error en sus enunciados; la búsqueda de un marco universal de exactas categorías e ideas (una *mathesis universalis*); la mecanización del pensamiento humano; la formulación, ilustración empírica y muestra de robustez frente intentos de falsación de modelos para la cognición; y, finalmente, en su encarnación más reciente, la gestación de la inteligencia computacional. A veces el proyecto se oculta, calla o disfraza su nombre para evitar la burla, por vergüenza o por miedo; otras, se infla ensimismado y orgulloso: excede lo razonable, bordea lo risible, causa rechazo y decepción, carcajada, desprecio. Entre tanto, los científicos que, con ilusión, nos hemos embarcado en esta aventura, avanzamos con persistencia, insistencia y tenacidad; casi con obstinación.

La primera dificultad del proyecto es definir su objeto y sus objetivos. La inteligencia es un concepto construido y, por tanto, mutable. Elude unívoca definición perdurable, precisa y certera. Basta recorrer la historia de lo que es considerado inteligente en las ciencias de lo humano (psicología o pedagogía) para ilustrar esta dificultad: las diferencias en tiempos y entre escuelas y el énfasis variable sobre lo cualitativo, lo cuantitativo, lo práctico, lo conceptual, lo holístico, lo creativo, lo emocional, etc. multiplican las imágenes de la inteligencia, como si el concepto fuera un espejismo encerrado en un caleidoscopio conceptual.

En las ciencias de la computación nos encontramos con que la fluidez del discurso transforma el concepto de inteligencia en una entelequia evanescente, huidiza: los grandes logros alcanzados en, sin afán de ser exhaustivo, la inferencia lógica (*General Problem Solver*, diseñado en 1959 por Herbert A. Simon, John C. Shaw, and Allen Newell) el magistral juego de ajedrez (*Deep Blu*e, creado en los años noventa del siglo pasado por un equipo de ingenieros de IBM), la experta conducción (el vehículo autónomo construido recientemente por un equipo de *Google* dirigido por Sebastian Thrun), la supremacía sobre el humano en juegos de preguntas y respuestas (el sistema de inteligencia artificial *Watson*, diseñado en IBM por un equipo liderado por David Ferrucci, que superó largamente a sus oponentes en el concurso de la televisión estadounidense *Jeopardy*), o la inducción automática a partir

de datos (aprendizaje máquina), tienen como consecuencia la gradual desaparición de dichas tareas de entre las que orgullosos incluimos en nuestro humano catálogo de actividades inteligentes.

Por consiguiente, la inteligencia artificial es, por inconfesada confabulación de los humanos, una meta inalcanzable. Frente a las ciencias de lo natural, como la física, la química o la biología, cuya muerte es anunciada, nunca agotaremos las preguntas en las ciencias de lo artificial, como la economía, la sociología o la informática, entendida esta última como ciencia del tratamiento -procesamiento, elaboración, aprovechamiento- automático de la información.

Con el fin de que nos sea posible hablar sin ambigüedades sobre del proyecto de la inteligencia artificial, y a pesar de que habrá quien se muestre reacio a aceptar tal premisa, convendremos en que del éxito en tal empresa consistirá en concebir, diseñar y, finalmente, construir un sistema cuyo comportamiento sea reconocido como inteligente por otros sistemas inteligentes; por ejemplo, demos muestra de generosidad, por los humanos. No es, sin embargo, mi objetivo tratar de banal la tarea o su alcance: Independientemente de si hemos construido inteligencia o si simplemente la hemos simulado con exquisita precisión, de forma que no nos sea posible diferenciarla de aquella que algunos distinguirían como real, alcanzar esta meta, aun siendo tan reducida su ambición, constituiría en sí un logro notable desde el punto de vista científico, tecnológico, cultural y social.

En la inteligencia artificial confluyen numerosos saberes y ciencias: la filosofía, la lógica, la matemática, la psicología, la economía y la informática, entre otras. El progreso hacia el esquivo objetivo de construir máquinas inteligentes requiere de numerosas ingenierías: la mecánica, la biomédica, las de la cognición y de la comunicación. En algún caso los artífices del proyecto han adoptado una visión antropomórfica: su meta es desentrañar y reconstruir la mente humana, incluyendo su impertinente apéndice, el cuerpo. Cuenta el taoísta *Tratado de la Perfecta Vacuidad* (*Lie Zi*) que Rey Mu del reino de Zhou contempló asombrado la figura humana del autómata que le ofreció como presente el ingeniero Yen Shih. La creatura caminaba con paso ágil, basculando la cabeza como lo haría un humano, cantaba con una perfecta afinación, e incluso cortejaba galantemente a las damas de palacio.

En la antigua Grecia y el antiguo Egipto, diestros ingenieros construyeron estatuas sagradas cuyas entrañas mecánicas, según creencia de los fieles, reproducían la naturaleza verdadera de los dioses, y, por tanto, estaban dotadas de emoción y aliento. Se cuenta también que en Praga a finales del siglo XVI el rabino Jehuda Loew ben Bezalel dio vida a un *Golem*, un ser artificial con forma toscamente humana, que, según la leyenda, salvó a los judíos de Praga de las persecuciones en su contra. Mary Shelley recreó en su novela "*Frankestein, o el moderno Prometeo*" la gestación de un nuevo hombre ilustrado, poseído de una aguda y dolorosa conciencia de sí mismo (y de su diferencia), poseedor de una delicada sensibilidad e inteligencia, radicalmente humano. El ingeniero Leonardo Torres y Quevedo construyó en 1915 el Autómata Ajedrecista, precursor de los actuales robots antropomorfos. Una de las últimas encarnaciones de nuestro soñado hermano mecánico es el robot humanoide *Kismet* (destino, suerte, o hado, de origen turco) diseñado por el equipo liderado por Cynthia L. Breazeal (1967-) en *Massachusetts Institute of Technology* (MIT). *Kismet* emula de manera convincente las expresiones faciales que son espejo de emociones en humanos y otros mamíferos.

Buscando la imitación, la inteligencia otros investigadores han renegado de los placeres y tentaciones de la carne y adoptan como objetivo explorar, entender y reconstruir la mente incorpórea: bien encarnada, con base orgánica (neurociencia) o inorgánica (cerebros electrónicos), o virtual (simulaciones de la dinámica redes de neuronas). Esta es una de las metas inconfesadas (¿inconfesables?) que persiguen el proyecto europeo *Human Brain Project* y el estadounidense *Brain Initiative*.

Finalmente, podemos abandonar todo referente humano y, en lugar de buscar reflejo o compañero, construir solucionadores de problemas, sistemas expertos, homeostáticos, adaptativos, de

planificación y de control, máquinas que aprenden, evolutivas, enjambres y colonias de agentes, que, como individuos, son simples pero que, como colectivo, muestran comportamientos inteligentes.

En el relato de esta joven empresa científica merecen distinción numerosos protagonistas, de los que haré, obedeciendo preferencias y afinidades personales, una selección, la cual en ningún caso debe ser tomada como resultado de la aplicación de un criterio sistemático y de académico rigor.

La complejidad del proyecto de la inteligencia artificial hace que las contribuciones, sobre todo las más recientes, sean en su mayor parte resultado de esfuerzos colectivos. No obstante, de entre los agentes individuales del proyecto sobresale Alan Turing, uno de los científicos más importantes del siglo XX. De haber sobrevivido a la agresión que sufrió por su preferencia sexual (quizás en una sociedad más abierta, más cuidadosa, más civilizada que la Inglaterra de mediados del siglo pasado; en suma, más humana) y a otros accidentes, habría cumplido 102 años en 2014. Turing fue un científico multifacético: matemático, lógico, pionero de la informática, formó parte de un equipo de ingenieros, científicos y técnicos que descifraron códigos utilizados por los alemanes durante la segunda guerra mundial. Hizo asimismo contribuciones seminales en el campo de la inteligencia artificial y de la morfogénesis. Posteriormente fue condenado judicialmente por ser homosexual. La pena consistió en su castración mediante un tratamiento con hormonas. Murió poco tiempo después de haber finalizado el tratamiento; probablemente por suicidio. A su memoria dedico este ensayo acerca de su proyecto imaginado. El relato se estructura en torno a su contribución. Pero para narrar la historia, narremos antes la prehistoria.

2. LA INTELIGENCIA ARTIFICIAL ANTES DE TURING (A. T.)

En esta etapa la historia del empeño se confunde con la del pensamiento humano; o, de manera más precisa, con la del descubrimiento, análisis y formalización del pensamiento racional. Recordemos que lo artificial no se opone a lo natural (por el contrario, lo artificial podría ser considerado como una sofisticada expresión de nuestra humana naturaleza), sino a lo no elaborado por el hombre. El motivo conductor que impregna la mayoría de las investigaciones de este periodo es la mecanización del razonamiento. Nuestra lista de desiderata para los atributos de esta mecanización incluye universalidad, corrección y completitud: Universalidad, para garantizar la validez general del método, independientemente del tema concreto sobre el que discurramos; corrección, con el fin de evitar en nuestro razonamiento falacia o error; completitud, para hacer posible razonar todo lo razonable. Comenzaremos como comienzan muchas de nuestras historias sobre la razón en Occidente: en torno a nuestro mar, un mar rodeado de tierras, el *Mare Nostrum:*

a) En las orillas del Mediterráneo

Nuestra historia tiene, como Dionisio, hijo de dos madres, un doble nacimiento:

En el extremo oriental del Mediterráneo Aristóteles hace por primera vez un tratamiento sistemático de los principios que gobiernan la inferencia correcta. Su gran contribución, el silogismo:

(i) Todos los entes que cumplen la relación X, cumplen la relación Y.
(ii) Todos los entes cumplen la relación Y, cumplen la relación Z.
(iii) Por lo tanto, los entes que cumplen la relación X, cumplen la relación Z.

En este modo de inferencia, si asumimos que, en nuestra particular circunstancia, las premisas (i) y (ii) son correctas, se sigue necesariamente que la conclusión (iii) también es correcta. Por ejemplo, podemos, para ser confrontados con nuestra naturaleza animal, hablar de las relaciones entre humanos (X), mamíferos (Y) y animales (Z), o de las que se dan entre filósofos (X), humanos (Y), mortales (Z), con el resultado de (re)descubrir la fugacidad de la vida de los que buscan la sabiduría. La corrección de este tipo de razonamiento es tan evidente que es aceptada sin necesidad de prueba:

constituye una “deducción perfecta”. Si transformáramos todo razonamiento en perfecta deducción, podríamos discurrir sin error: el razonamiento, codificado como manipulación sintáctica, es liberado de la falible y ambigua semántica. Pero, si hemos eliminado del razonamiento el significado, ¿cómo podemos afirmar que se trata de un discurso sobre el mundo? La realizad aparece tanto al inicio, cuando suponemos que, en la situación concreta acerca de la que se discurre, las premisas son verdades pertinentes, como en el punto final del razonamiento, cuando la conclusión es interpretada como un enunciado acerca de dicha particular situación. Pero, ¿no podrían deslizarse el error y la falacia por estas imperfectas juntas entre la deducción y lo real? Deberemos aceptar esta fractura en nuestro, de no ser por ella, perfecto esquema de mecanización del pensamiento.

Más de mil años más tarde, en la isla de Mallorca, Ramón LLull (1235-1315) fabulaba sobre qué método idear para convencer mediante razonamiento a musulmanes, judíos y escépticos de lo erróneo de sus creencias y de lo verdadero de la cristiana fe: la existencia de Dios, la necesaria encarnación de Dios en Cristo, el Dios uno y trino y otros dogmas de compleja argumentación. Los métodos utilizados habitualmente para la conversión de infieles apelaban a un principio de autoridad, bien textual (la Biblia), bien personal (el Papa). El problema que Llull advertía en estos argumentos era que los infieles contaban con autoridades propias, en ocasiones superiores en civilización, en intelecto a las cristianas: el Corán y los mulás para musulmanes; el Talmud y los rabinos para judíos. No obstante, las tres religiones comparten creencias, ciencia y filosofía, y semejante cultura, valores y moral. Por ello, sería posible acordar unas verdades elementales, que no necesitarían demostración por ser en sí evidentes.

A modo de ejemplo, acerca de la bondad, “*Bonea es raó per la qual bo fa bé, e per qui bona cosa es esser e mala cosa es no esser*”. Dado que una combinación de estas verdades simples no puede sino producir una verdad también válida, más compleja, ¿por qué no proceder a enumerar todas las combinaciones posibles e interpretar cada una de ellas para descubrir, explicar o demostrar nuevo conocimiento? Para llevar a cabo estas combinaciones, Llull diseñó artefactos mecánicos y grafos jerarquizados. Uno de ellos consiste en una ensambladura de coronas circulares concéntricas de distinto diámetro, divididas en sectores circulares, cada uno de los cuales era etiquetado con el nombre de una noción perteneciente a una determinada clase de conceptos. Las coronas podían ser giradas de manera independiente, de forma que, al leer cada una de las combinaciones de símbolos radialemente, se obtenía una verdad compuesta a partir de los bloques de verdades elementales. El *Ars Generalis Magna*, incomprendido por la mayoría, despreciado por no constituir un sistema lógico formal, proscrito por la Iglesia largo tiempo y finalmente ignorado, anticipa, no obstante, la importancia del razonamiento combinatorio en la inteligencia artificial.

b) El andamiaje del razonamiento deductivo

Los esfuerzos de Llull no tuvieron continuidad en ningún discípulo o escuela filosófica de su época. La idea de que el razonamiento es simplemente una forma de computación reaparece en Europa en la filosofía de Thomas Hobbes (1588-1679). En el primer capítulo de *De Corpore* se lee: “*Por razonamiento, entiendo computación. Y computar es tomar la suma de muchas cosas realizada de manera simultánea o conocer lo que resta cuando una cosa ha sido tomada de otra. Razonar es por lo tanto lo mismo que sumar o sustraer*” [1]. Gottfried Wilhelm Leibniz (1646-1716) advierte de que la única manera de resolver disputas es hacer concreto el razonamiento, como en matemáticas: Calculemos para saber quién lleva razón. Para ello, ideó un lenguaje para representar simbólicamente conceptos e ideas (*characteristica universalis*), y un sistema o un dispositivo de deducción lógica mediante manipulación tipográfica de los símbolos de dicho lenguaje (*calculus ratiocinator*).

Este diseño haría posible razonar con ideas de la misma forma que la manipulación de números en una máquina de calcular permite razonar con cantidades. Siguiendo el trazado establecido por Leibniz, George Boole (1815-1864) formula la lógica formal como un álgebra en *Mathematical Analysis of Logic* y el posterior *An Investigation of the Laws of Thought on Which are Founded the*

Mathematical Theories of Logic and Probabilities. En el álgebra de Boole las variables denotan conceptos que pueden tomar valores de verdad. Por ejemplo, la variable *A*, que hace referencia a la proposición "Dios existe" puede tomar únicamente el valor *A = Verdadero*, en el caso de que, efectivamente, Dios existiera, o *A = Falso*, si fuéramos huérfanos de Dios. Este átomo simbólico con valor de verdad podría ser combinado con otros átomos mediante conectores lógicos para formar proposiciones compuestas, de significado más complejo. Por ejemplo, la regla condicional NOT(A) $\Rightarrow$ B significaría "Si Dios no existe, el hombre está solo", con la denotación obvia para la variable booleana B. De nuevo, ¿a qué lugar ha sido relegado el mundo? La respuesta no es nueva: únicamente mediante un misterioso proceso de experiencia directa en el mundo físico podría el humano determinar el valor de verdad de dichos átomos simbólicos (¿y la máquina?, ¿podría ella también?, ¿podemos, mediante experiencia, determinar el valor de verdad de A?). El comienzo del siglo XX vio, con las formulaciones de la lógica de predicados realizadas por Frege y Pierce, el apogeo del proyecto de construir (o descubrir) un cálculo de la razón. No había en ese momento indicio alguno del futuro descalabro.

c) La mecanización del pensamiento

Dado que, si creyéramos a Hobbes, Leibniz y Boole, el pensamiento humano no es más que una forma de cálculo automático, ¿por qué no construir máquinas capaces de realizar dicha tarea? Los trabajos del matemático, filósofo e ingeniero Charles Babbage (1791-1871) y de la matemática y escritora Ada Lovelace (1815-1852) son un paso más en la encarnación del pensamiento racional en la máquina. Babbage diseñó, pero no llegó a construir, un artefacto mecánico programable que incluía en su lenguaje de programación mecanismos de control de flujo (estructuras condicionales), y estructuras de repetición (bucles). En lenguaje moderno, la máquina es Turing-completa, ya que puede llevar a cabo cualquier cómputo que pueda realizar una máquina universal de Turing, la cual será descrita más tarde. Se anticipaba así más de 100 años a las primeras computadoras de propósito general programables fabricadas en 1940 bajo la dirección de Conrad Zuse (1910-1995). La máquina analítica de Babbage supone un avance cualitativo respecto a las calculadoras mecánicas desarrolladas por Wilhelm Shickard (1592-1635), Blaise Pascal (1623-1662) y Leibniz y a la máquina de diferencias diseñada por el propio Babbage. Ada Lovelace es considerada como la primera programadora de la historia. Inspirada por los cartones perforados en los que se codificaban los patrones de las telas tejidas en los telares mecánicos enunció: "la máquina analítica teje patrones algebraicos del mismo modo que el telar de Jacquard teje flores y hojas". Más de cien años después, millones de tarjetas perforadas serían procesadas por computadoras electrónicas haciendo realidad esta hermosa metáfora.

d) Ascenso y caída del edificio de la lógica

A principios del siglo XX, matemáticos prominentes se propusieron esclarecer los fundamentos de las Matemáticas. David Hilbert (1862-1943) aborda la denominada crisis fundacional de las Matemáticas mediante un programa de formalización en la línea esbozada 200 años antes por Leibniz: Establecer un lenguaje en el que formular proposiciones matemáticas sin ambigüedad y definir reglas precisas para manipular correctamente dichas proposiciones, de forma que proposiciones nuevas correctas puedan ser derivadas de manera mecánica de las anteriores, suponiendo que estas son también correctas. Utilicemos lógica para descubrir verdades sobre la propia lógica: Así como el razonamiento (X) fue reducido al cálculo (Y), el cálculo debería ser reducido a la lógica (Z). En consecuencia, todo razonamiento (X) sería simplemente manipulación de proposiciones lógicas (Z).

La lógica proposicional desarrollada por Boole fue extendida por, entre otros, Gottlob Frege (1848-1925) y Charles Peirce (1839-1914), mediante la incorporación de variables que podían representar colecciones de objetos o hacer referencia a objetos indeterminados, cuantificadores (universal: "Para todo X,…", existencial: "Existe al menos un Y, tal que…") y predicados (relaciones). Alfred North Whitehead (1861-1947) y Bertrand Russell (1872-1970) abordaron en *Principia Mathematica* la hercúlea tarea de reducir las Matemáticas a la Lógica: un juego formal, en el que, a

partir de proposiciones indemostradas que se suponen correctas (axiomas), se pueden derivar otras proposiciones (teoremas) también correctas, aplicando la maquinaria de inferencia de la lógica simbólica.

Sin embargo, el programa de Hilbert resultó ser quimera: El matemático Kurt Gödel (1906-1978) demuestra en su teorema de incompletitud que no es posible construir un sistema formal axiomático que, teniendo suficiente capacidad expresiva como para describir la aritmética de los naturales, sea a la vez correcto y completo. Es decir, si el sistema es coherente y no genera proposiciones contradictorias, existirán necesariamente teoremas correctos que no es posible derivar a partir de los axiomas mediante un procedimiento constructivo. Es más, no es posible determinar si para una proposición dada, candidata a ser teorema (es decir correcta, si correctos fuera los axiomas), existe dicha prueba constructiva o no. El teorema de Gödel tiene un eco devastador en nuestros dispositivos de cómputo programables: De acuerdo con el teorema demostrado por Alan Turing en 1936 relativo al *problema de la parada*, no es posible diseñar un procedimiento infalible que nos permita determinar, para todos los posibles programas, operando sobre cualesquiera datos de entrada, si la ejecución de un programa concreto sobre ciertos datos de entrada, terminará su ejecución en un tiempo finito, generando una respuesta, o si, por el contrario, continuará calculando por un tiempo indefinido sin detenerse. ¿Si es mecánica, está la inteligencia humana también sujeta a estas limitaciones fundamentales? Y si lo estuviere, ¿supondría esto una demostración de que no es posible la absoluta inteligencia?

3. ALAN TURING

Alan Mathison Turing (1912-1954) es el personaje central (no los hay secundarios o principales) en nuestra historia. Literalmente, ocupa el centro del devenir de la búsqueda de inteligencia en nuestras criaturas y artefactos: culmina (y de alguna forma agota) la vía de la mecanización del razonamiento deductivo e inicia el camino azaroso de la inteligencia artificial imperfecta, falible, aproximada, heurística, centrada en el significado, encarnada en el mundo de los hombres, no en el de las ideas. Las contribuciones de Turing son numerosas y, en su mayoría originales y sorprendentes.

El hito que transforma la computación en Ciencia es la ideación de una máquina de cómputo universal por parte de Turing. La máquina es tan sencilla que ni siquiera necesita ser construida: podemos simular su funcionamiento. Opera sobre una cinta de papel que funciona como dispositivo de entrada y salida de datos. La cinta está dividida en celdas. Cada una de estas celdas contiene uno de dos símbolos, convencionalmente bits (dígitos binarios: ceros o unos). Un reloj, cual metrónomo, pauta el progreso del procesamiento. La máquina puede encontrarse en uno de una colección finita de estados. En cada ciclo de reloj la máquina lee el símbolo que está inscrito en una única celda de la cinta. Dependiendo del símbolo leído y del estado en el que se encuentre, borra el símbolo leído, escribe uno nuevo, no necesariamente distinto, y se desplaza a una celda adyacente, bien a la derecha, bien a la izquierda de la actual. También puede alcanzar un estado terminal, que señalaría la finalización del cómputo; el resultado de la operación del programa sobre los datos de entrada, inicialmente escritos en la cinta de papel, queda registrado también en ella.

Turing conjetura, junto con Alonzo Church (1903-1995), que lo computable es lo que puede ser computado en tal máquina. La conjetura de Church y Turing aplica a (aunque quizás fuera más preciso decir que define su función) todos los dispositivos de cómputo: desde nuestros dispositivos portátiles, teléfonos y ordenadores, hasta los más potentes supercomputadores. La simplicidad y potencia de esta máquina permite demostrar teoremas sobre qué es computar, cuáles son sus posibilidades, sus limitaciones y sus límites.

En 1950 Turing publica el artículo "*Computing machinery and intelligence*", que prefigura gran parte del desarrollo de la inteligencia computacional de la segunda mitad del siglo XX: máquinas que

razonan, que evolucionan y se adaptan, que adquieren y generan conocimiento; máquinas que aprenden. También en dicho artículo se describe el Test de Turing, en el que se propone la definición funcional de la inteligencia que hemos convenido utilizar en el inicio de este ensayo. Este test enfrenta a un interrogador humano, a un sistema computacional, y a otro humano en interacción verbal remota, de forma que no haya sesgos relacionados con el aspecto externo del adversario. El objetivo del interrogador es desenmascarar a la máquina tras una conversación de una duración razonable. Si el interrogador humano se muestra incapaz de distinguir humano y máquina, el sistema habrá superado el test y será acogido en la comunidad de seres inteligentes.

4. LA INTELIGENCIA ARTIFICIAL DESPUÉS DE TURING (D. T.)

El florecimiento de la inteligencia artificial en la segunda mitad del siglo XX, en marcado contraste con el formalismo de su prehistoria, se caracteriza por su pragmatismo: Se usan heurísticas, razonamientos aproximados, algoritmos falibles que, si bien no garantizan encontrar la respuesta óptima -a veces, ni siquiera garantizan encontrar respuesta-, obtienen, en una gran proporción de los casos de interés práctico, resultados suficientes.

a) El nacimiento de una disciplina.

El acta de nacimiento de la inteligencia artificial es la propuesta elaborada en 1955 por los investigadores John McCarthy (1927-2011), Marvin L. Minsky (1927), Nathaniel Rochester (1919-2001) y Claude E. Shannon (1916-2001) de realizar un estudio con 10 personas durante 2 meses sobre el diseño y la construcción de máquinas que piensan. El punto de partida es la conjetura de que todos los elementos y procesos de la inteligencia pueden ser representados de manera precisa. Por ello, son susceptibles de ser simulados con la ayuda de máquinas. En la propuesta se plantea abordar, entre otros, los siguientes problemas: el lenguaje natural, la simulación de redes de neuronas, la mejora autónoma de sistemas, la manipulación computacional de abstracciones, sin olvidar la aleatoriedad y la creatividad. Todo ello en un caluroso verano en Nuevo Hampshire.

b) Retrato del ordenador como idiot savant

El éxito de los sistemas computacionales que desempeñan la función de los expertos humanos en un área de saber específica fue uno de los hitos tempranos de la inteligencia artificial, que hacían presagiar un brillante futuro para ella: El sistema DENDRAL, desarrollado en Stanford por un grupo de investigadores que incluía a Edward Feigenbaum (1936), Bruce Buchanan (1940-), Joshua Lederberg (1925-2008), y Carl Djerassi (1951), era capaz de identificar compuestos químicos (espectrometría de masas) a partir de la masa de los fragmentos generados tras ser sometida la muestra a un bombardeo con electrones. Pasado el deslumbramiento inicial, los sistemas expertos pasaron a ser meras herramientas, por su estrecho foco (el sistema no es útil fuera del área de especialización para la que fue diseñado), rigidez (el sistema no cambia, no se adapta, no aprende) y falta de sofisticación (MYCIN contiene cientos de reglas condicionales, excesivas para que un humano pueda manejarlas, insuficientes para representar el mundo en toda su complejidad y esplendor). Más recientemente, los esfuerzos se han centrado en demostrar competencia en mundos más complejos. Sin embargo, desde las contundentes victorias de la máquina *Deep Blue* en ajedrez, y del sistema computacional *Watson* en el concurso de preguntas y respuestas de la televisión estadounidense *Jeopardy*, los humanos hemos perdido interés en estos juegos: nos sentimos incómodos en la derrota (¡en nuestro humano terreno!, ¡por parte de máquinas!.) En la actualidad estos expertos sin conciencia están siendo entrenados para convertirse en nuestros futuros cuidadores médicos.

c) Lenguajes, autómatas formales y computación

Entres los esfuerzos por identificar las estructuras formales del pensamiento, destaca la formidable contribución de Noam Chomsky (1928): Por medio de un refinado análisis de las

estructuras que hacen posible la comunicación humana Chomsky describe mecanismos sintácticos subyacentes a todo lenguaje, tanto el de la máquina como el humano. Los distintos mecanismos de reconocimiento y producción de lenguaje se enmarcan en una jerarquía de gramáticas de creciente capacidad expresiva. La jerarquía de lenguajes se corresponde con una jerarquía paralela de máquinas: los autómatas formales, que tienen capacidad de resolver distintos tipos de problemas. Del mismo modo que, gracias a Turing, sabemos que existen problemas para cuya solución no podemos construir máquina alguna, habrá lenguajes para los que no existen autómatas que hagan posible su procesamiento.

d) Imitatio naturae

Imitar los mecanismos y estrategias perfeccionados por procesos naturales de generación de variabilidad y selección es una de las estrategias que nos conducen, por atajo, a recrear la inteligencia. En concreto, las redes de neuronas artificiales, la computación evolutiva (por ejemplo, los algoritmos genéticos), o la simulación de colectivos con comportamiento inteligente (inteligencia de enjambre) son paradigmas que, utilizados con sabiduría y prudencia, han permitido abordar problemas cuya resolución, diríamos la mayoría, requiere inteligencia. Centrémonos en el cerebro humano y su correlato artificial: las redes neuronales.

Santiago Ramón y Cajal (1852-1934) es probablemente el científico español más influyente (un inexplicable pudor me impide utilizar la palabra principal) de la historia. Por medio de técnicas de tintura desarrolladas por Camillo Golgi (1843-1926), las cuales permitían hacer visible la estructura de los tejidos mediante la fijación de cromato de plata en la membrana de las células, Ramón y Cajal realizó hermosos dibujos de neuronas, de los que se puede disfrutar el Instituto Cajal del Consejo Superior de Investigaciones Científicas. Gracias este cuidadoso trabajo de laboratorio llegó a la conclusión de que el sistema nervioso estaba formado por una red de neuronas. Curiosamente Golgi, con quien Ramón y Cajal compartió el premio Nobel en 1906, estaba convencido de la teoría opuesta; es decir, de que el sistema nervioso era una única red conexa.

Partiendo de estas y otras investigaciones histológicas el neurofisiólogo Warren S. McCulloch (1898-1969) y el lógico y matemático autodidacta (y, de acuerdo con Wikipedia, suicida cognitivo) Walter Pitts (1923-1969) proponen en un artículo, cuyo intrigante y levemente ominoso título: "*A Logical Calculus of Ideas Immanent in Nervous Activity*" (1943) desvela la ambición del proyecto, un modelo electrónico de neurona. En 1969, el partero y veterano pionero de la IA, Marvin L. Minsky (1927-), junto con Seymour Papert (1928), aventajado discípulo del psicólogo Jean Piaget (1896-1980), publicaron "*Perceptrons: An introduction to computational geometry*", uno de los libros clave en la inteligencia artificial conexionista, que, paradójicamente, al abordar las limitaciones de los perceptrones de una sola capa en la resolución de problemas de clasificación no separables linealmente, contribuyó a la parálisis de las investigaciones sobre redes neuronales durante casi dos décadas, coincidentes con el periodo conocido como el invierno de la inteligencia artificial.

En 1986 las investigaciones sobre redes neuronales recuperaron ilusión e impulso cuando David E. Rumelhart (1942-2011), Geoffrey E. Hinton (1947-) y Ronald J. Williams redescubrieron un método de entrenamiento para perceptrones multicapa por propagación de los errores hacia atrás en la red (*backpropagation*), el cual ya había sido analizado por Paul J. Werbos (1947-) en su tesis de 1974. Similares ciclos de expectativas desbordadas y grandes decepciones se repetiría en los años posteriores. En la actualidad nos encontramos en un periodo de euforia motivado por la posibilidad de ajustar los pesos de las sinapsis que median la comunicación entre neuronas en redes profundas (compuestas por varias capas ocultas) y por el inusitado y oportunista interés por los grandes volúmenes de datos (*big data*). Preparémonos para sobrellevar con ecuanimidad el vértigo del ascenso, la crueldad de la caída.

e) Máquinas que aprenden

El problema de cómo obtener la base de conocimiento que podría ser después codificada (por ejemplo, como una colección de axiomas) para después ser sometida a manipulaciones (por ejemplo, utilizando mecanismos de inferencia lógica) para obtener conocimiento (por ejemplo, un nuevo teorema) fue abordado de manera tardía en la inteligencia artificial. A diferencia de la deducción, que es un problema formalmente bien definido (¿mediante qué procedimiento puedo derivar conclusiones correctas a partir de conocimiento que supongo correcto?), el razonamiento inductivo tiene pies de barro: ¿Cómo puedo tener la certeza de que la hipótesis que formulo a partir de los datos generaliza y es aplicable a datos del mismo tipo que aún no he visto? Si la cantidad de datos por examinar fuera ilimitada (como habitualmente es el caso), existe siempre la posibilidad de que la hipótesis sea falsada: El conocimiento adquirido es, por tanto, provisional, sin esperanza de llegar a ser consolidado, independientemente de la evidencia, siempre finita, que coleccionemos. Gradualmente, sin embargo, el aprendizaje automático ha llegado a ocupar un lugar central en la inteligencia artificial. Dentro de este campo, destacaré a Leslie Valiant (1949-), quién nos enseñó qué se puede y qué no se puede aprender, Leo Breiman (1928-2005), por sus contribuciones a la inducción automática de árboles de decisión y de conjuntos de clasificadores (*bagging*, bosques aleatorios, etc.) y a Vladimir Vapnik (1936-) por construir el marco formal en el que analizar de manera rigurosa el problema de inducción a partir de datos.

Pendientes de ser contadas como merecen quedan las historias de otros grandes científicos que han contribuido al proyecto de la reencarnación de la inteligencia, esta vez en la máquina. Mencionemos únicamente dos: Norbert Wiener (1894-1964) y su monumental "*Cybernetics: Control and Communication in the Animal and the Machine*", y el premio Nobel de Economía Herbert A. Simon (1916-2001), con su ensayo sobre un nuevo tipo de ciencia (que requiere de un nuevo método científico) "*The Sciences of the Artificial*".

Tampoco hemos prestado suficiente atención a dudas, críticas y críticos. Expongamos a modo de reparación el punto de vista del filósofo John R. Searle (1932), quien, si bien acepta la posibilidad de que lleguemos a construir artefactos que reproduzcan comportamientos inteligentes con asombrosa fidelidad, afirma que estos serán siempre carentes de la llama que define la inteligencia verdadera: la intencionalidad. Para ilustrar su punto de vista, nos invita a realizar el experimento imaginado (*Gedankenexperiment*) de la habitación china: Supongamos que somos encerrados en una habitación en la que disponemos de lápiz, papel y un manual de instrucciones para la manipulación de ideogramas. Nuestra tarea es recoger los mensajes escritos en chino en una hoja de papel que alguien nos desliza por una rendija desde fuera de la habitación, y responder con otro mensaje, que hemos de componer siguiendo las reglas que se encuentran en el manual. Las reglas son sencillas: si encuentras una cierta combinación de ideogramas en una posición dada, escribe esta otra combinación de ideogramas.

El manual es tan completo y bien diseñado que los mensajes compuestos tienen sentido. Desde el punto de vista de nuestra desconocida interlocutora, estamos llevando una conversación en mandarín perfectamente cuerda, dado que, aunque no seamos conscientes de ello, nuestro mensaje contiene réplicas creíbles a los textos que ella nos hace llegar. Es decir, nuestro sistema, la habitación compuesta por el libro y por nosotros mismos, pasaría un test análogo al de Turing, ya que hemos simulado de manera convincente entender chino. No obstante, si planteamos la pregunta de dónde reside dicho entendimiento, no encontramos lugar para ello: el libro, por ser objeto inanimado, no puede ser su sede; nosotros mismos somos conscientes de que no entendemos chino. Luego habremos de concluir que, dado que puede ser localizado, el entendimiento debe estar ausente. Aunque puede que, si bien sus partes carezcan de él, el sistema completo posea dicho atributo: es decir, el entendimiento podría estar deslocalizado.

Para concluir, hablemos de futuros utópicos y distópicos: ¿debiéramos haber mantenido cerrada la mente de Pandora? A este respecto se han alzado las voces de prominentes científicos (Stephen

Hawking, Stuart Russell, Max Tegmark, Frank Wilczek en *The Independent*, 2014/05/01) advirtiendo que, si bien el éxito en la creación de un sistema artificial dotado de inteligencia constituiría el evento de mayor envergadura de la historia de la humanidad, podría ser también el último, a menos que tomemos en serio la importancia del proyecto y sus implicaciones. Apartaremos por el momento este pensamiento, como se hace con los malos sueños (de la razón): habiendo sido de Pandora la mente abierta, son ya los males y bienes que en ella tengan su origen parte inseparable e irrenunciable de nuestra naturaleza, que incluye lo artificial y la inteligencia. Esta nueva exploración merece la pena, a pesar de que el conocimiento al que nos lleve duela, desgarre y nos fuerce a elegir; es decir, a seguir pensando, a seguir viviendo.

Para conocer más sobre este tema:

1. Anthony Bonner (1997): "*What was Llull up to?*". Lecture Notes in Computer Science, Vol. 1231, pp. 1-14. [http://dx.doi.org/10.1007/3-540-63010-4_1]
2. Brian Randell (1982): "*From Analytical Engine to Electronic Digital Computer: The Contributions of Ludgate, Torres, and Bush*". IEEE Annals of the History of Computing, Vol. 4, nº 4, pp. 327-341, October-December. [http://dx.doi.org/10.1109/MAHC.1982.10042]
3. Alan M. Turing (1950): "*Computing machinery and intelligence*". Mind, Vol. 59, nº 236, pp. 433-460.
4. Javier De Felipe (2005): "*Cajal y sus dibujos: Ciencia y arte*". Arte y Neurología: 213-230.
5. John McCarthy, Marvin L. Minsky, N. Rochester y Claude E. Shannon: "*A proposal for the Dartmouth Summer Research project on Artificial Intelligence*" [http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html]
6. Norbert Wiener (1965): "*Cybernetics or Control and Communication in the Animal and the Machine*". The MIT Press.
7. Marvin Minsky (1988): "*The Society of Mind*". Simon & Schuster.
8. Herbert A. Simon (1996): "*The Sciences of the Artificial*". The MIT Press.
9. "*The MIT Encyclopedia of the Cognitive Sciences (MITECS)*", editada por Robert A. Wilson and Frank C. Keil, The MIT Press (2001)
10. Nils J. Nilsson (2010): "*The quest for Artificial Intelligence: A History of Ideas and Achievements*" Cambridge University Press. [http://www.cambridge.org/us/0521122937]

Nota: Para confeccionar este trabajo se han consultado éstas y otras fuentes, las cuales, por numerosas, que no por carecer de interés o importancia, no han sido citadas. Mencionar que entre ellas se encuentra Wikipedia, una enciclopedia en línea, a menudo denostada, pero que, en numerosas entradas, proporciona gran riqueza de contenidos (en amplitud y profundidad), precisión y fidelidad a las fuentes originales. De ella se han extraído fechas e información de base. Para la documentación sobre personalidades y conceptos de la filosofía ha sido extremadamente útil el recurso a la enciclopedia *The Stanford Encyclopedia of Philosophy*, editada en formato electrónico por Edward N. Zalta [http://plato.stanford.edu/].