

# Adónde nos lleva la Inteligencia Artificial?

Joseph C. Várilly

Escuela de Matemática, Universidad de Costa Rica, 11501 San José, Costa Rica

Revista Desarrollo 5 (1987), 81–86

## Resumen

La Inteligencia Artificial es una realidad que ya existe y se proyecta hacia el futuro. La pregunta *¿podrán pensar las máquinas?* se ha lanzado muchas veces antes de 1950, cuando Alan Turing tuvo al osadía de contestarla positivamente. Un cuarto de siglo después, los avances logrados dieron lugar a una nueva inquietud: *¿deberían pensar las máquinas?* Hoy en día tenemos una breve oportunidad de elegir entre dos opciones. ¿Debemos proseguir construyendo esta maravilla científica, o debemos detenernos antes de que sea tarde?

## I

Para explorar esta pregunta, empezaré con una reseña de algunas obras que abordan este problema desde diferentes puntos de vista. Norbert Wiener escribió en 1960 un artículo [1] titulado *Algunas consecuencias morales y técnicas de la automatización*, el cual es en parte una sinopsis del penúltimo capítulo de su famoso libro *Cibernética* [2]. Recibió una réplica de Arthur Samuel [3], inventor de un programa temprano para el juego de damas. Pamela McCorduck ha escrito una historia del primer cuarto de siglo de Inteligencia Artificial, sus logros y sus controversias [4]. En un breve artículo reciente [5], William Barrett comenta las ideas de Turing; su ensayo está tomado de un libro [6] sobre el desarrollo del cientismo moderno.

### El fantasma del Golem

La posibilidad de crear un ser inteligente no humano es un viejo sueño. Desde las máquinas de Dédalo hasta los robots de Karel Čapek, la confección de seres artificiales es un tema recurrente de la mitología. El caso más interesante es quizás el del Golem, una criatura semiinteligente supuestamente creado por el rabino Judah Loew en el siglo XVI, usando fórmulas de la Cábala. La criatura, inicialmente dócil, luego se rebela y hace estragos antes de que el rabino lo logra agarrar y lo destruye [7]. El mismo *leitmotif* recorre la novela de Mary Shelley sobre la criatura de Victor Frankenstein, y el cuento de Goethe sobre el aprendiz del brujo. Wiener hace referencia explícita a este último cuento [2, p. 176] para explicar sus temores sobre las consecuencias de crear máquinas sofisticadas: él sostiene que no se puede anticipar todas las actuaciones de una máquina con cierto

grado de autonomía. Si una máquina lograra trascender las limitaciones de sus diseñadores y actuara con juicio propio, sería altamente peligrosa. Samuel contesta que las máquinas no pueden poseer originalidad y que hay – y siempre habrá – una laguna completa entre la elaboración por una máquina de los deseos del hombre y el desarrollo de una voluntad propia.

Otro orden de fantasías sobre las máquinas inteligentes aparece en la ciencia ficción moderna, con sus androides y robots. Isaac Asimov en 1950 [8] postuló la creación en el siglo XXI de un “cerebro positrónico” que incorpora de modo fundamental sus tres *leyes de robótica*: estas leyes prohíben al robot causar daño o desobedecer a los seres humanos, obviando así el problema del Golem y dando base a ficciones fascinantes. Pero no pasan de ser ficciones, pues no hay una mínima indicación de cómo un artefacto pueda incorporar estas “leyes de robótica”. Una idea más honesta aparece en una novela futurista de Robert Heinlein en 1966 [9]: un ordenador gigantesco, encargado de todo el manejo de una colonia lunar, un día alcanza el grado de complejidad necesario para llegar a ser autoconsciente, y se “despierta”. El agnóstico Heinlein no tiene reservas en suponer que la conciencia es solamente una función de la complejidad material del cerebro (negando así la hipótesis de Samuel); aceptada esta suposición, Heinlein delinea en esta y otras novelas el desarrollo de personalidad en sus seres mecánicos.

## **El simulacro del poeta**

Los robots pensantes son – todavía al menos – fantasías. Turing fue el primero en abordar la pregunta *¿podrán pensar las máquinas?* para explorar la practicalidad de responder *que sí* [10]. Turing enumera varias objeciones a la posibilidad de máquinas pensantes: que la inteligencia es función del alma humana; que el teorema de Gödel lo prohíbe; que una máquina no puede ser autoconsciente; que una máquina nunca podrá hacer diversas cosas (tales como distinguir lo bueno de lo malo, enamorarse, aprender de la experiencia, tener sentido de humor, etc.); que una máquina no puede crear nada original; que el comportamiento humano no puede reducirse a un conjunto de reglas de decisión. A todos estos argumentos Turing ofrece refutaciones (algunas convincentes, otras no). Explora en detalle la posibilidad de que una máquina aprenda de sus experiencias, comenzando con una simulación de una mente infantil. Pero la parte más osada de su artículo es su consideración de una máquina que escribe poesía. Ofrece un diálogo entra un *Testigo* que ha escrito un poema y un *Interrogador* que trata de averiguar si el Testigo fue el autor original; el Interrogador no sabe que el Testigo es una máquina, pues no están frente a frente, y las respuestas no permiten descubrirlo.

William Barrett, un intelectual neoyorquino de prestigio y durante muchos años editor de la *Partisan Review*, contesta esta afirmación de Turing [5]. Primero, observa que Turing defiende un poema ya escrito pero no indica cómo se puede producir el poema en primer lugar. Luego va al grano: *¿en qué consiste escribir un poema, que va más allá de congregar una colección de signos y símbolos?* Barrett señala que quien lee poesía no lee los poemas uno por uno, sino que el lector serio lee buena parte de la obra de un poeta, porque esos poemas viven por la personalidad del poeta. Además, el poeta vive en el presente, pero a través de su poesía expresa la herencia del pasado y de sus antecesores, y transmite su propio sentido histórico. También, un poeta cambia y envejece con el tiempo e ilumina su poesía con las penas y alegrías de su maduración. Barrett resume con la pregunta: *¿cuánto de todo esto puede hacer una máquina inorgánica e históricamente aislada?*

## Puntos de controversia

Los optimistas responden: *¿quién sabe?* Su punto de vista está bien presentada en el libro de Pamela McCorduck [4], audazmente titulado *Las Máquinas Quienes Piensan*. Ese libro es una sinopsis, para no especialistas, del desarrollo de la Inteligencia Artificial (IA) desde 1950 a 1975. Aparecen Minsky, Newell, Simon, Papert, McCarthy y muchos otros que han hecho contribuciones fundamentales a la disciplina de la IA. La autora nos muestra cómo las ideas nebulosas iniciales sobre máquinas que aprenden fueron concretadas en mecanismos y programas específicos, tales como jugadores de damas y ajedrez, interpretadores de idiomas, analizadoras visuales, sistemas expertos, y aplicaciones de la IA al aprendizaje humano. El catálogo de logros es impresionante y justifica la confianza de los trabajadores en el campo cuando afirman que se podrá eventualmente aproximarse mucho más a un artefacto inteligente. El subtítulo del libro es: *Una indagación personal en la historia y perspectivas de la inteligencia artificial*. La parte histórica es fascinante, pues revela los esfuerzos, logros, reveses y sorpresas de un gran trabajo científico; pero aún más importante son sus perspectivas.

Todavía no se ha creado una máquina inteligente y autónoma, y mucha gente opina que nunca se creará. Los escépticos abundan. El ejemplo más conocido es el “caso Dreyfus”. Hubert Dreyfus, un filósofo de Berkeley, ha atacado con vehemencia la idea de razonamiento artificial, en un libro *Lo que los computadores no pueden hacer* [11]. Dreyfus señala que la hipótesis gobernante de la investigación sobre la IA, de que la actividad mental que llamamos inteligencia pueda ser descrito mediante un conjunto de reglas que gobiernan la conducta intelectual y moral, y que la creación de nueva inteligencia equivale a descubrir e implementar esas reglas, es parte de un punto de vista filosófico particular y reduccionista. (Barrett también observe la singularidad del fundamento filosófico de los partidarios de una máquina inteligente, y los remonta a los empiricistas británicos). Dreyfus cometió el error de desdeñar los sistemas de jugar ajedrez, afirmando (en 1965) que un niño de diez años podría ganar a cualquier máquina en ajedrez; unos años más tarde, él se enfrentó con un nuevo programa sofisticado que lo venció contundentemente.

Pamela McCorduck clasifica los argumentos en contra de la IA en cuatro tipos: argumentos de emoción; de dificultades insuperables; de carencia de ejemplos existentes; y de consideraciones éticas. La carencia de ejemplos es un argumento débil: ninguna máquina jamás ha escrito un poema decente, pero en el futuro ¿podrá hacerlo? Para este problema Barrett señala dificultades insuperables, pero ¿quién puede garantizar que son en verdad invencibles? Samuel afirma que una máquina nunca podrá adquirir voluntad propia (una opinión que compartió von Neumann); Wiener no estaba tan seguro. De todos modos, la historia hasta ahora ha mostrado que la frase *nunca podrá* es de empleo delicado, como aprendió el pobre Dreyfus.

Estipulemos, entonces, que sea posible crear una máquina o artefacto que exhibe un comportamiento que merece el nombre de inteligencia. En este caso debemos enfrentar el argumento más importante de todos: *¿debería creárselo?* Joseph Weizenbaum, en su libro de 1976, *Computer Power and Human Reason* [12], argumenta que poder no implica deber. Junto con Kenneth Colby, el ingeniero Weizenbaum había diseñado en 1965 un programa de conversación interactiva entre máquina y ser humano, llamado ELIZA (por la Bella Dama que aprendió a hablar cada vez mejor). Colby, un psiquiatra, lo usó para consultas clínicas con resultados asombrosos: los respondientes “conversaron” con la máquina como si fuera un ser humano, aun sabiendo que se trataba de un programa automático. Weizenbaum concluye que hay dominios afectivos en la vida humana donde

una máquina no debe entremeterse. Al inicio de su libro escribe [12]:

*Un punto principal de este libro es precisamente que nosotros, todos, hemos hecho excesivamente al mundo en un computador, y que este rehacer del mundo en la imagen del computador empezó mucho antes de la venida de los computadores electrónicos.*

## II

### **Tecnología y voluntad humana**

Los problemas éticos que surgen con la nueva tecnología de la inteligencia artificial trascienden esta tecnología porque son, como apunta Weizenbaum, problemas de definición humana. Wiener se asustó porque podríamos crear un Golem que escapa de nuestro control y quizás nos destruye; pero una amenaza más real e inmediato es la de desvalorizar nuestra humanidad al nivel de una máquina y así destruirnos nosotros mismos. ¿Podremos obtener los beneficios de la tecnología de la IA sin caer en ninguno de estos dos peligros?

El miedo a la presencia de una máquina versátil en nuestras vidas es tal vez sólo una quimera. Wiener lanzó su aviso sobre la “automatización” (automation), un vocablo de los años cincuentas. En aquel tiempo esa palabra significaba el desplazamiento de mano de obra por máquinas (no automáticas) en la agricultura e industria [13]. Treinta años después, podemos apreciar que esa fue una etapa transitoria y que hoy en día la mecanización genera tanto o más empleo de lo que desplaza. Las sociedades que más empleos generan son precisamente aquellas que son tecnológicamente más avanzadas. Eso es un pobre consuelo para aquellas cuyas habilidades se vuelven redundantes, pero indica que la cura reside en adaptar la nueva tecnología a las necesidades actuales y no en rehuirla. Los japoneses ya usan robots programables en la producción de vehículos en lugar de hacer todo con brazos humanos; el resultado neto de esto da un beneficio económico a la sociedad japonesa.

Claro es que las máquinas y robots actuales no son “inteligentes”, pero el recelo ante su empleo y la oposición a la inteligencia artificial surgen de la misma fuente. Nadie quiere confiar su destino a la decisión de un proceso automático. Se intuye que una tal decisión no es menos arbitraria que una decisión humana, y que es mucho menos apelable. Ya en 1960 Wiener presintió una realidad cotidiana de nuestros días: la política de confiar una respuesta ante la agresión nuclear a programas que no sólo coordinen el lanzamiento de misiles sino que sean capaces de decidir en pocos minutos si una respuesta es necesaria. Lo que Wiener tal vez no previó es que la llamada política de Destrucción Asegurada Mutua sería defendido tenazmente ante intentos de adoptar una estrategia más flexible con los mismos argumentos que se ensayan contra la inteligencia artificial: dificultades insuperables, falta de ejemplos existentes, emociones y argumentos de carácter ético.

El abuso de la tecnología se plantea agudamente cuando consideramos su uso en la guerra. Nuestro siglo ha visto las guerras más atroces de la historia, agudizadas por la tecnología bélica, y esto da lugar a preguntar si la tecnología es un mal en sí. Algunas armas, como el gas venenoso, son absolutamente inexcusables; otras, como el tanque, el símbolo moderno del poder ofensivo, tienen un papel ambiguo: el tanque fue introducido en 1916 como un arma cuasi defensiva para romper el dominio mortífero de la ametralladora en las trincheras de Francia. Con un poco de reflexión, se ve que el frío no está en las cobijas: lo que hace horroroso un arma es la voluntad humana de

emplearla. El gas venenoso no fue usado en la segunda guerra mundial (entre combatientes, al menos); y después de Nagasaki, nadie todavía se ha atrevido a usar otra vez la bomba atómica. El peligro de la guerra nuclear es que algún día se atreverá a usar las armas sin medir las consecuencias.

En un discurso reciente, Solzhenitsyn nos recuerda: *Las fallas de la conciencia humana, desprovisto de su dimensión divina, han sido un factor determinante en todos los grandes crímenes de este siglo. El primero de estos fue la primera guerra mundial, y gran parte de nuestro apuro actual proviene de allí* [14]. Lo que a veces se olvida es que esta guerra fue tan vasta y calamitosa debido a que al principio sus generales entregaron su juicio a la decisión de un computador. Recordemos las palabras de Weizenbaum: mucho antes de la venida de los computadores electrónicos, habían otros artefactos con la misma imagen; en este caso, los horarios de los trenes de movilización para llevar a quince millones de soldados a los frentes de guerra. Los generales austriacos, rusos y alemanes empezaron sus preparativos de movilización antes de las declaraciones de guerra en agosto de 1914, pero los franceses e ingleses quedaron indecisos y durante dos días no entraron en la guerra. El Kaiser alemán vio la oportunidad y ordenó a su jefe de Estado Mayor, el general von Moltke, a cambiar el plan y marchar contra Rusia solamente. *Imposible!* exclamó von Moltke, *el plan es automático, no se puede cambiar*. El día siguiente los alemanes invadieron a Bélgica [15].

El temor de que una máquina nos quite la iniciativa y nos lleve a situaciones calamitosas no es cosa del futuro; por una falta de imaginación moral, ya lo hemos vivido.

### **¿Hacia dónde vamos?**

El problema da perder el control de nuestro destino, no se debe entonces a la presencia o ausencia de máquinas inteligentes, sino al comportamiento intelectual y ético de los seres humanos. Nuestro siglo xx se ha caracterizado por el abuso de las ciencias pero eso no es culpa de las ciencias, sino de la proclividad de los humanos a abusarlas. Por más inteligente que sea un mecanismo o programa que algún día se creara, la responsabilidad para sus acciones no sería propia, sino que quedaría sobre los hombros de sus creadores. (Esa es la parábola de Victor Frankenstein.)

Pero como hay posibilidades y ejemplos de abuso, también hay oportunidades para el buen uso de los computadores. Wiener temía que el computador llegaría a ser esclavizante, pero no es necesariamente así. Lewis Thomas ha dicho: *La ciencia es útil. Es indispensable. Pero cuando avanza, lo hace al producir una sorpresa, y no podemos especificar de antemano cuál tipo de sorpresa nos gustaría* [16]. Una sorpresa de los últimos veinte años ha sido el efecto liberalizante, y no esclavizante, de los computadores. En lugar de una concentración de poder e información en menos de estados y grandes corporaciones que usarían gigantescas “mainframes” para controlar cada aspecto de nuestras vidas, ahora vemos la información dispersada en millones de pequeños microcomputadores, abriendo un espacio de autonomía personal. Sobre el futuro nada está dicho, nada está escrito. Debemos prevenir los abusos, pero también tendremos posibilidades de emplear las nuevas tecnologías para el bien. Lo que necesitamos para ello es la fe en el futuro; y eso es algo que ninguna máquina, por tonta o inteligente que sea, nos puede dar.

Me abstengo de elegir un lado en el debate sobre si es posible o no construir un mecanismo inteligente. Simpatizo con Barrett cuando carga contra el reduccionismo de los partidarios de la IA y declara que hay una esfera reservada a la acción humana. Pero no es claro todavía cuáles son los límites de esa esfera. Es inclusive posible que tengan razón quienes afirman que es factible construir

un artefacto que exhiba aspectos de inteligencia tales como capacidad de decisión, comprensión de nuevos mensajes, e inclusive generación de conceptos originales. Claro es que esto no cubre todo lo que llamamos inteligencia – Lewis Thomas [16] duda que un computador podrá sentir sorpresa, por ejemplo – pero abre perspectivas impresionantes. Enfrentemos entonces la posibilidad de que algún día, quizás no muy remoto, los computadores dejarán de ser meros instrumentos de trabajo y serán capaces de aprender y de enseñar.

El beneficio más importante que la IA ya nos ha dado es su aplicación al aprendizaje humano. Para construir una máquina que aprende, hay que comprender el cómo del aprendizaje humano. Aun si la IA fracasara en sus proyectos mas ambiciosos, la contribución que ya ha hecho a la teoría del aprendizaje infantil sería una justificación suficiente.

Samuel concluye diciendo: *Las invenciones del hombre son instrumentalidades que pueden ser usados por santos y pecadores* [3]. Debemos rechazar la falacia materialista de que la posesión de tales instrumentalidades limita nuestra voluntad. Tenemos responsabilidades éticas que no podemos abdicar a la decisión de una máquina, como hizo von Moltke: esta parte de nuestra humanidad es inalienable. Pero si mantenemos la claridad sobre quienes somos y qué somos los humanos, no tenemos nada que temer del desarrollo de la inteligencia artificial. Al final de cuentas, al motor del avance científico no es la abundancia de los recursos materiales, sino la fe en el futuro y nuestra esperanza ante el porvenir.

## Referencias

- [1] Norbert Wiener, “Some moral and technical consequences of automation”, *Science* **131** (1960), 1355–1358.
- [2] Norbert Wiener, *Cybernetics*, MIT Press, Cambridge, MA, 1961.
- [3] Arthur L. Samuel, “Some moral and technical consequences of automation: a refutation”, *Science* **132** (1960), 741–742.
- [4] Pamela McCorduck, *Machines Who Think*, W. H. Freeman, San Francisco, 1979.
- [5] William Barrett, “Why computers can’t be poets”, *Commentary* **81** (1986), 63–65.
- [6] William Barrett, *Death of the Soul: from Descartes to the Computer*, Anchor/Doubleday, New York, 1986.
- [7] Véase, por ejemplo: Jorge Luis Borges y Margarita Guerrero, *El libro de los seres imaginarios*, Kier, Buenos Aires, 1967.
- [8] Isaac Asimov, *I, Robot*, Gnome Press, New York, 1950.
- [9] Robert Anson Heinlein, *The Moon is a Harsh Mistress*, Berkley, New York, 1966.
- [10] Alan Turing, “Computing machinery and intelligence”, *Mind* **59** (1950), 433-460. Traducido como: “Maquinaria computacional e inteligencia”, en *Mentes y Máquinas*, Editorial de la UNAM, México, 1970.
- [11] Hubert Dreyfus, *What Computers Can’t Do: A Critique of Artificial Reason*, Harper & Row, New York, 1972.
- [12] Joseph Weizenbaum, *Computer Power and Human Reason*, W. H. Freeman, San Francisco, 1976.
- [13] Véase, por ejemplo: Samuel Eliot Morison, *The Oxford History of the American People*, tomo 3, Oxford University Press, Oxford, 1965.

- [14] Aleksandr Solzhenitsyn, Discurso de aceptación del Premio Templeton por Progreso en la Religión, Londres, 1983. Disponible en línea:  
<https://www.templetonprize.org/laureate-sub/solzhenitsyn-acceptance-speech/>
- [15] Basil Liddell-Hart, *History of the First World War*, Pan, Londres, 1970.
- [16] Lewis Thomas, "The value of basic science", *Rochester Review* **44** (1981), 1–7.