

# La inteligencia artificial

Vicenç Torra

V. Torra, (2011), La inteligencia artificial, Revista Lychnos, Fundación General del CSIC, Diciembre de 2011.

**Resumen.** Este artículo describe la inteligencia artificial y sus áreas principales. Se describen también algunas de las aplicaciones más relevantes desde un punto de vista científico. El artículo fue publicado en la revista Lychnos de la Fundación General del CSIC en 2011.

La inteligencia artificial (IA) es una de las ramas de la informática, con fuertes raíces en otras áreas como la lógica y las ciencias cognitivas. Como veremos a continuación, existen muchas definiciones de lo que es la inteligencia artificial. Sin embargo, todas ellas coinciden en la necesidad de validar el trabajo mediante programas. H. A. Simon, uno de los padres de la IA, nos sirve de ejemplo pues afirmó en un artículo en 1995 que “el momento de la verdad es un programa en ejecución<sup>1</sup>”. Las definiciones difieren en las características o propiedades que estos programas deben satisfacer. Comentaremos las diferencias más adelante, en la sección primera.

La inteligencia artificial nace en una reunión celebrada en el verano de 1956 en Dartmouth (Estados Unidos) en la que participaron los que más tarde han sido los investigadores principales del área. Para la preparación de la reunión, J. McCarthy, M. Minsky, N. Rochester y C. E. Shannon, redactaron una propuesta en la que aparece por primera vez el término “inteligencia artificial”. Parece ser que este nombre se dio a instancias de J. McCarthy.

## 1. ¿Qué es la inteligencia artificial?

La propuesta citada más arriba de la reunión organizada por J. McCarthy y sus colegas incluye la que puede considerarse como la primera definición de inteligencia artificial. El documento define el problema de la inteligencia artificial como aquel de construir una máquina que se comporte de manera que si el mismo comportamiento lo realizara un ser humano, éste sería llamado inteligente<sup>2</sup>.

Por tanto, esta primera definición de inteligencia artificial se focaliza en el comportamiento humano. Esta no es, sin embargo, la única aproximación existente en la actualidad pues algunos autores han mostrado que esta definición limita los objetivos del campo. La crítica fundamental es que el comportamiento humano no debe ser la única inspiración en la construcción de programas ni tampoco la única medida para evaluar el rendimiento de los programas. Esta mayor apertura<sup>3</sup> permite considerar cuatro

---

<sup>1</sup> “The moment of truth is a running program” (p.96) en: H. A. Simon, Artificial Intelligence: an empirical science, Artificial Intelligence 77 (1995) 95-127.

<sup>2</sup> “For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving”. J. McCarthy, M. L. Minsky, N. Rochester, C. E. Shannon, A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955.

<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>

<sup>3</sup> La consideración de estas dos dimensiones aparece en el texto de Stuart Russell y Peter Norvig, Artificial Intelligence: A Modern Approach. del que han aparecido varias ediciones desde de 1995. La presentación de la inteligencia artificial en el libro prima la opción (d): actuar racionalmente.

definiciones, construidas a partir de dos dimensiones: cual es el objetivo de los programas y cual es la forma de evaluarlos. Así, encontramos las cuatro definiciones siguientes:

- a) **Actuar como las personas.** Esta definición corresponde a la de McCarthy y, por tanto, el modelo a seguir para la evaluación de los programas corresponde al comportamiento humano. Esto es, un programa actúa correctamente si tiene un comportamiento similar al que tienen las personas. El llamado Test de Turing (1950) también utiliza este punto de vista. El sistema Eliza, un bot (programa software) conversacional es un ejemplo de ello<sup>4</sup>.
- b) **Razonar como las personas.** En la definición anterior el foco está en el resultado de la conducta, y no en como se consigue esa conducta. Por tanto, un procedimiento formalizado matemáticamente pero que no corresponde a la forma de razonamiento humana es válido para cumplir el test de Turing si los resultados del procedimiento son similares a los de las personas. La alternativa propuesta aquí es desarrollar sistemas que razonen del mismo modo que las personas. La ciencia cognitiva<sup>5</sup> utiliza este punto de vista.
- c) **Razonar racionalmente.** En este caso, la definición también se focaliza en el razonamiento, pero aquí se parte de la premisa de que existe una forma racional de razonar. La lógica permite la formalización del razonamiento y se utiliza para este objetivo. Como es sabido, el razonamiento lógico no tiene porqué coincidir siempre con el razonamiento que utilizan las personas.
- d) **Actuar racionalmente.** El objetivo no es como se consiguen los resultados sino los resultados que se consiguen. La forma de razonamiento deja de ser importante y se subraya la necesidad de que las decisiones y las actuaciones que se realicen sean las correctas. Sin embargo, a diferencia de la primera definición y el test de Turing, la evaluación de la corrección no se hace en base a una persona, que puede equivocarse, sino en base a medidas que son objetivas. Por ejemplo, el objetivo de un programa en un juego como el ajedrez será ganar, y si es posible jugar mejor que cualquier otro jugador. Para cumplir este objetivo puede ser indiferente si las decisiones se toman con mecanismos similares a las personas o tan otros mecanismos.

**El efecto ELIZA.** El bot conversacional ELIZA desarrollado por el Prof. J. Weizenbaum del MIT en el año 1966 imitaba a un psicólogo. Su implementación, desde un punto de vista computacional, era muy sencilla y sus respuestas se basaban en unas pocas reglas. Sin embargo, el proceso interactivo con los usuarios provocaba respuestas emocionales importantes. El efecto ELIZA es la tendencia a observar el comportamiento de un ordenador como análogo al de los humanos.

---

<sup>4</sup>ChatGPT puede verse también como Eliza como un sistema conversacional que intenta emular el comportamiento humano. En este sentido es interesante el trabajo de Cameron R. Jones and Benjamin K. Bergen (2024) Does GPT-4 pass the Turing test?, <https://arxiv.org/abs/2310.20216>. El trabajo analiza hasta que punto ChatGPT-4 pasa el test de Turing, y que en la mejor configuración lo pasa el 49.7%.

<sup>5</sup> Desde un punto de vista de la computación, la inteligencia artificial y la robótica es interesante el libro de David Vernon (2014) Artificial cognitive systems: A primer, MIT Press.

Además de las definiciones mencionadas más arriba hay aún otra clasificación de la inteligencia artificial, según cuales son los objetivos finales de la investigación en este campo. Son la inteligencia artificial fuerte y la débil.

- 1) **Inteligencia artificial débil.** Se considera que los ordenadores únicamente pueden simular que razonan, y únicamente pueden actuar de forma inteligente. Los partidarios de la inteligencia artificial débil consideran que no será nunca posible construir ordenadores conscientes, y que un programa es una simulación de un proceso cognitivo pero no un proceso cognitivo en si mismo.
- 2) **Inteligencia artificial fuerte.** En este caso se considera que un ordenador puede tener una mente y unos estados mentales, y que, por tanto, un día será posible construir un ordenador con todas las capacidades de la mente humana. Este ordenador será capaz de razonar, imaginar, etc.

## 2. Temas en la inteligencia artificial

Aunque existen puntos de vista diferentes sobre que es la inteligencia artificial, hay un acuerdo importante sobre cuales son los resultados atribuibles a esta rama de la informática, así como a la clasificación de los métodos y técnicas desarrollados. Repasamos a continuación los cuatro grandes temas de la inteligencia artificial.

- a) **Resolución de problemas y búsqueda.** La inteligencia artificial tiene como objetivo resolver problemas de índole muy diferente. Para poder cumplir este objetivo, dado un problema es necesario formalizarlo de manera que un programa pueda resolverlo. Problemas muy diferentes pueden expresarse de forma similar de acuerdo con la idea de que el objetivo del programa es pasar de un estado inicial (la situación actual) a un estado final u objetivo (la situación deseada) a partir de la aplicación de un conjunto de acciones. Para cada problema concreto es necesario definir de forma clara cuales son los estados posibles, la situación inicial, a que situación deseamos llegar y cuales son las acciones posibles. Entonces una solución es encontrar la secuencia de acciones necesarias para pasar del estado inicial a un estado final que satisfaga nuestros objetivos. Por ejemplo, si queremos integrar una función, el estado inicial es la expresión a integrar, las acciones posibles corresponden a las reglas de integración conocidas y el objetivo es obtener una expresión sin el símbolo integral. Una solución es una secuencia de acciones que nos permite pasar de la expresión a integrar a una expresión sin símbolo integral.
- b) **Representación del conocimiento y sistemas basados en el conocimiento.** Es frecuente que los programas en inteligencia artificial necesiten incorporar conocimiento del dominio de aplicación para poder resolver los problemas. Por ejemplo, un sistema de diagnóstico médico deberá tener almacenados los conceptos médicos y las relaciones entre ellos. La lógica<sup>6</sup> es uno de los mecanismos de representación del conocimiento, y la inferencia lógica permite deducir nuevos hechos a partir del conocimiento del sistema. Existen, sin embargo, otros formalismos de representación<sup>7</sup>, cada uno con su forma de

---

<sup>6</sup> Existen diferentes tipos de lógica que se utilizan para problemas diferentes. Por ejemplo, lógica proposicional, lógica de predicados, lógica modal, lógica difusa, lógica temporal.

<sup>7</sup> Se utiliza el término Ontología para denotar la representación de los conceptos, datos, entidades y las relaciones entre ellos, que se necesitan para construir un sistema inteligente.

representar el conocimiento y sus mecanismos para deducir nuevos hechos a partir de lo que ya es conocido.

- c) **Aprendizaje automático**<sup>8</sup>. El rendimiento de un programa puede incrementarse si el programa aprende de la actividad realizada y de sus propios errores. Se han desarrollado métodos con este objetivo. Un ejemplo muy sencillo es el siguiente: en un sistema para integrar expresiones el sistema puede aprender (recordar) las expresiones ya integradas anteriormente para reutilizar el resultado más tarde. Existen también herramientas que permiten extraer conocimiento a partir de bases de datos. Por ejemplo, herramientas que extraen expresiones lógicas a partir de bases de datos. Las redes neuronales y el aprendizaje profundo (*deep learning*) son también ejemplos de aprendizaje de datos. En este caso aprenden un modelo de comportamiento a partir de los datos disponibles.
- d) **Inteligencia artificial distribuida**. Durante sus primeros años la inteligencia artificial era monolítica. Sin embargo, el incremento en el número de procesadores de los ordenadores y sobretodo la existencia y ubicuidad de la red han incrementado el interés en soluciones distribuidas. Las soluciones van desde versiones paralelas de métodos ya existentes a nuevos problemas relacionados con los agentes autónomos. Estos últimos son programas software con autonomía suficiente para tomar decisiones que interaccionan con otros programas, y que pueden incluso colaborar entre ellos para resolver un problema.

**Ejemplo de sistema de aprendizaje.** Dado un fichero con los registros de 4 variables booleanas  $v_1$ ,  $v_2$ ,  $v_3$  y  $v_4$  queremos obtener una expresión lógica que relacione  $v_4$  en términos de las otras tres.

$v_1$	$v_2$	$v_3$	$v_4$
1	1	1	1
1	1	0	1
1	0	1	1
0	0	0	0
0	1	0	0

Un sistema de aprendizaje podría encontrar la relación  $v_4 = v_1$ , pero también otras relaciones como la relación  $v_4 = v_1 \wedge (v_2 \vee v_3)$  a partir de estos datos.

Además de los cuatro temas mencionados más arriba, existen otros que están fuertemente relacionados con la inteligencia artificial. De hecho, algunos autores como Winston<sup>9</sup> los consideran parte fundamental del área. Estos temas son los enumerados a continuación.

<sup>8</sup> El aprendizaje automático se conoce en inglés por *Machine Learning*. Es importante subrayar que el aprendizaje automático es únicamente una de las áreas en las que se divide la inteligencia artificial. Existen aplicaciones de inteligencia artificial que se limitan a utilizar machine learning. Son sistemas basados en datos. Este es el caso de las redes neuronales y el aprendizaje profundo.

<sup>9</sup> Sometimes, you hear the phrase core AI used by people who regard language, vision, and robotics to be somehow separable from the mainstream of artificial intelligence. However, in light of the way progress on language, vision, and robotics can and has influenced work on reasoning, any such separation seems misadvised. Winston (1994) Artificial Intelligence.

- a) El lenguaje natural.
- b) La visión artificial.
- c) La robótica.
- d) El reconocimiento del habla.

Como se verá a continuación muchas de las aplicaciones más llamativas utilizan algunas de las técnicas relacionadas con estos temas.

### 3. Algunas aplicaciones

Hasta la fecha se han desarrollado muchas aplicaciones que utilizan algunos de los métodos o algoritmos diseñados en el área de la inteligencia artificial. En esta sección repasamos algunas de las aplicaciones existentes más vistosas o que han tenido relevancia histórica. Sin embargo, estas no son las únicas aplicaciones existentes, pues hay métodos desarrollados en esta rama de la informática que se utilizan en aparatos de uso cotidiano o en el software utilizado por empresas y corporaciones. Por ejemplo, por una parte encontramos los algoritmos de búsqueda citados más arriba en los sistemas que construyen horarios teniendo en cuenta las restricciones de las entidades e individuos que participan en ellos. Por otra parte, los métodos de aprendizaje se usan para recomendar productos en las tiendas virtuales y para seleccionar los anuncios que se nos proporcionan al visitar ciertas páginas web. Otro ejemplo es el de los sistemas difusos, uno de los métodos de representación del conocimiento, que han sido aplicados con éxito en problemas de control de muy diversa índole. Existen tanto cámaras digitales como lavadoras que incorporan en su interior un sistema difuso.

Pasamos a ver a continuación algunas de las aplicaciones más relevantes.

**1. Aplicaciones en los juegos.** Durante décadas la inteligencia artificial se ha dedicado a los juegos con el fin de derrotar a los mejores jugadores humanos. Se ha considerado siempre que la habilidad de jugar era propia de la inteligencia. Las damas y el othello fueron derrotados primero. En el año 1997 fue el turno del ajedrez. Ahora prácticamente queda únicamente el Go por batir.

- **Las damas.** Chinook, un programa desarrollado desde el año 1989 por el equipo de Jonathan Schaeffer de la Universidad de Alberta, se convirtió en el año 1994 en el campeón mundial de damas. El programa incluye una base de datos con aperturas de los mejores jugadores y otra de situaciones finales cuando quedan 8 o menos fichas en el tablero. El mismo equipo investigador demostró años más tarde, en el año 2007, que cuando se juega a las damas de forma perfecta ningún jugador puede ganar. Esto significa que una estrategia óptima por parte de los dos jugadores únicamente puede llevar a tablas. La dificultad de esta demostración estriba en que pueden haber aproximadamente  $500 \cdot 10^{20}$  tableros posibles, o, en palabras, 50,000 trillones de tableros. El problema es un millón de veces más difícil que el 4 en raya. La información sobre este programa se puede encontrar en: <http://webdocs.cs.ualberta.ca/~chinook/>.
- **El ajedrez.** Durante muchos años se desarrollaron inventos y programas con el fin de ganar en el juego del ajedrez, un juego considerado por todos como muy difícil. Charles Babbage (1791-1871), Leonardo Torres Quevedo (1852- 1936) y Claude Shannon (1916-2001) son tres personajes conocidos que dedicaron su

tiempo a este juego. Charles Babbage, conocido por su máquina analítica precursora de los ordenadores modernos, fue uno de los primeros en considerar este juego, Torres Quevedo desarrolló en el año 1910 un autómatas que jugaba rey y torre contra rey y Claude Shannon publicó en el año 1950 el primer artículo de como implementar un programa de ajedrez. Sin embargo, no fue hasta mayo del año 1997 cuando Deep Blue venció al campeón humano G. Kasparov en Nueva York. El programa desarrollado por IBM utilizaba hardware específico, bases de datos que permitían al programa jugar de forma perfecta en las situaciones finales con 7 o menos fichas en el tablero, y algoritmos de búsqueda del tipo minimax para encontrar la mejor solución en todos los otros casos. Se puede encontrar información sobre Deep Blue en la página web <https://www.ibm.com/history/deep-blue>

- **El Go.** Mientras los otros juegos ya han sido derrotados, no existe en la actualidad ningún programa informático que tenga un nivel suficiente en Go como para ganar a los buenos jugadores humanos. Hace años que se considera el Go un juego mucho más difícil que el ajedrez. La dificultad estriba en las dimensiones del tablero (son 19 x 19, con 361 intersecciones), el número de movimientos posibles en cada tablero y la dificultad de definir funciones que evalúen correctamente un tablero dado. Actualmente se han conseguido algunos programas que tienen un buen nivel en un tablero reducido de 9 x 9. Los programas que tienen un buen rendimiento no utilizan el mismo algoritmo de búsqueda que el ajedrez (el mínimax) sino el UCT.

**2. Aplicaciones en robótica.** Las aplicaciones en robótica se han desarrollado desde el principio de la informática con diversidad de objetivos: la automatización de procesos industriales, las aplicaciones militares, y la exploración espacial. Mientras los primeros robots estaban orientados a realizar actividades repetitivas, actualmente se busca una mayor autonomía en relación a su capacidad de tomar decisiones. La evolución de la robótica también ha pasado por su intento de construir robots con forma humana, y con capacidad de andar. A continuación se enumeran algunos de los logros más importantes en esta área.

- **Los robots mascotas.** Existen en la actualidad muchos robots como diversión o juegos. Podemos subrayar dos de ellos. Los japoneses Paro y el Aibo. El primero es un robot terapéutico para reducir el estrés de los pacientes, y que permite incrementar su socialización. Aibo, presentado en el año 1999 y implementado y comercializado por SONY, es un robot en forma de perro con un sistema de visión y programable.
- **Los robots de exploración y reconocimiento.** Se enviaron a la superficie de Marte dos robots el 3 y el 24 de enero de 2004, respectivamente. Son los robots Spirit y Opportunity. Spirit estuvo activo hasta el año 2010 y Opportunity aun está en funcionamiento. El 26 de noviembre de 2011 se ha mandado un tercer robot, el Curiosity, que debe llegar a Marte en agosto de 2012. Este robot tiene navegación automática media de 30 m/hora (y máximade 90 m/hora). Se espera que su misión dure dos años.
- **Los robots bípedos.** P3 fue el primer robot bípedo capaz de caminar. Su construcción, realizada por la compañía Honda, acabó en el año 1997. En

octubre del año 2000, la misma compañía presentó el robot ASIMO (de Advanced Step in Innovative MObility). Este robot es la culminación de la serie de robots P, todos desarrollados con el objetivo de disponer de robots con forma y capacidades motrices humanas. ASIMO es un robot bípedo que puede andar y correr. Además puede reconocer objetos móviles, posturas y gestos, a partir de la información suministrada por sus cámaras.

**3. Aplicaciones en vehículos inteligentes.** Se han construido muchos tipos de vehículos con diferentes grados de autonomía. En el apartado anterior ya se han mencionado algunos robots. Aquí se mencionan vehículos que pueden llevar pasajeros.

- **El metro de la ciudad japonesa de Sendai (仙台地下鉄).** Este metro, desarrollado en el año 1987, fue el primero en el mundo con una conducción totalmente automatizada. Actualmente son muchas las líneas de metro en el mundo sin conductor. El sistema fue realizado bajo la dirección de Seiji Yasunobu, del laboratorio de desarrollo de sistemas de Hitachi, y basado en técnicas de lógica difusa.
- **Coches autónomos.** Stanley fue el coche ganador de la carrera “2005 DARPA Grand Challenge”. El coche autónomo y sin conductor completó el recorrido de 212.4 km en el desierto de Mojave en los Estados Unidos en 6 horas y 54 minutos. En el 2007 se realizó la “2007 DARPA Grand Challenge” que consistía en 96 km en área urbana (en la base de las fuerzas aéreas George en California). Los coches circulaban procesando en tiempo real las reglas de circulación del estado de California. El 20 de septiembre de 2011 el coche MadeInGermany de la Universidad Libre de Berlín<sup>10</sup> se desplazó por las calles de Berlín en un viaje de 80 km. El recorrido fue entre el Centro de congresos internacional y la puerta de Brandenburgo en el centro de Berlín.
- **Los vehículos aéreos no tripulados (UAV, Unmanned Aerial Vehicle).** El vehículo aéreo no tripulado Global Hawk fue el primero en cruzar el océano Pacífico sin paradas. Realizó el trayecto de Estados Unidos (California) a Australia en abril del año 2001. Sin embargo, aun necesita un piloto en una estación terrestre y otros operadores para analizar los datos. De acuerdo con Weiss (2011)<sup>11</sup> el principal problema de estos sistemas autónomos es que aunque pueden recoger muchos datos, les falta todavía capacidad de proceso para tratar los datos en tiempo real y actuar de forma inteligente a estos datos.

#### 4. Conclusiones

La informática ha avanzado enormemente desde sus inicios hace 70 años. La potencia de cálculo se ha estado multiplicando por dos cada 18 meses, siguiendo la Ley de Moore. Se cree que si la ley de Moore se sigue cumpliendo, para el año 2030 la capacidad de cálculo de un procesador corresponderá a la de una persona.

---

<sup>10</sup> El automóvil es completamente autónomo, aunque algunas informaciones, como la velocidad del recorrido, le vienen dadas y no son recogidas por sus cámaras. El coche reconoce la presencia de peatones, y los semáforos. Para información del coche alemán puede consultarse la siguiente página web: <http://autonomos.inf.fu-berlin.de/>

<sup>11</sup> Weiss, L. G. (2011) Autonomous robots in the fog of war, Spectrum August 2011, 27-31.

A su vez, la cantidad de información almacenada digitalmente en la actualidad es enorme. Los buscadores como Google tienen almacenadas copias de millones de las páginas web existentes, y los servicios de correo de las compañías tienen almacenados millones de nuestros mensajes. Las redes sociales registran cuáles son nuestros intereses y nuestras amistades. Las compañías almacenan cualquier información por insignificante que pueda ser, por si en el futuro les puede ser de alguna utilidad.

Naturalmente un aumento de la velocidad de computación y una mayor capacidad de almacenaje implicará que los sistemas dispongan de más recursos para tomar decisiones, y que estas decisiones se realicen de manera más informada, y a su vez de manera más personalizada.