Inteligencia artificial

Manuel Alfonseca
Modo de citar:
Alfonseca, Manuel. 2016. "Inteligencia artificial". En Diccionario Interdisciplinar Austral, editado por Claudia E. Vanney, Ignacio Silva y Juan F. Franck. URL=http://dia.austral.edu.ar/Inteligencia_artificial

Desde el principio de la historia de la informática se planteó el problema de si será posible algún día diseñar y programar computadores y sistemas complejos semejantes para que actúen con una inteligencia comparable, o incluso superior a la humana. Se trata de una cuestión que nos afecta profundamente desde el punto de vista sentimental, lo que explica el éxito de obras clásicas de ficción como *Frankenstein* o *El golem*, que expresan en forma literaria el viejo sueño de construir hombres artificiales, ya sea por medios biológicos no naturales, o en forma de autómatas mecánicos.

La discrepancia entre estas expectativas y el carácter repetitivo y rígido de la mayor parte de nuestras aplicaciones informáticas (incluso muchas de las más avanzadas en la actualidad) no ha sido suficiente para poner en duda la factibilidad de ese sueño. La verdad es que, a lo largo del tiempo, hay bastantes actividades humanas supuestamente inteligentes (como jugar al ajedrez) que han ido cayendo bajo el dominio de los computadores. Pero por una reacción comprensible, suele ocurrir que esas actividades, una vez dominadas por programas informáticos, ya no nos parecen tan inteligentes. Existe incluso una definición de *inteligencia artificial* basada en ello, que podemos considerar *irónica*, pues su intención no es seria, pero que encierra un fondo de verdad: *inteligencia artificial* sería **todo aquello que todavía no sabemos hacer** con un ordenador**. Como el horizonte, la *inteligencia artificial* de verdad parece alejarse de nosotros a medida que nos movemos hacia ella.**

Esta entrada pretende ofrecer al lector una visión actualizada de la inteligencia artificial, en sus dos acepciones de *inteligencia artificial débil* (la que ya tenemos y está a nuestro alcance) y de *inteligencia artificial fuerte* (la construcción y programación de máquinas tan inteligentes como nosotros o más), junto con una revisión de las diversas posturas filosóficas respecto a la cuestión de si este objetivo será factible, o por el contrario no será posible conseguirlo.

1 Definición de la inteligencia artificial ↑

La inteligencia artificial es una de las ramas más antiguas de la investigación en programación de ordenadores. Casi desde el principio de su historia, se han construido programas que se comportan de un modo que, cuando los seres humanos hacen lo mismo, solemos calificar de inteligente. Sin embargo, los avances en esta área no han sido constantes, pues se han visto sometidos a altibajos. Además, los investigadores no siempre se ponen de acuerdo en la definición de esta rama de la informática, por lo que no es fácil distinguir de forma clara y unívoca las disciplinas y aplicaciones que pertenecen a este campo.

Algunas de las definiciones del campo de la inteligencia artificial son claramente insatisfactorias. Así John McCarthy, que acuñó el nombre de la disciplina, la define así: *ciencia e ingeniería para la fabricación de máquinas inteligentes, especialmente programas de ordenador inteligentes* (McCarthy 2007). El problema de las definiciones de este tipo es que no se define lo que quiere decir *inteligente*.

Una definición parecida, bastante extendida, es esta: *estudio y diseño de agentes inteligentes* (Nilsson 1998, Legg 2007), donde un *agente inteligente* se define como *un ente que percibe su entorno y realiza acciones que le permiten alcanzar sus objetivos con las máximas posibilidades de éxito*. Aquí el problema es que no se define lo que significa el *éxito en alcanzar sus objetivos*.

La siguiente definición de la inteligencia artificial está bastante aceptada, al menos para lo que se ha hecho hasta ahora en este campo:

Definición 1: *Llamamos* inteligencia artificial *al conjunto de técnicas que tratan de resolver problemas relacionados con el proceso de información simbólica, utilizando para ello métodos heurísticos*.

No es nuevo que la información contenida en la memoria de un ordenador puede ser simbólica, pero en las aplicaciones de inteligencia artificial, en las que intentamos (en principio) que la máquina razone de forma parecida al hombre, la información que se procesa será a menudo equiparable a ideas o conocimientos y sólo podrá representarse de forma simbólica. Esto no significa que no pueda haber información numérica en una aplicación de inteligencia artificial. Por el contrario, aparece frecuentemente. Pero así como en la programación clásica puede ocurrir que toda o casi toda la información procesada por una aplicación determinada sea numérica, en inteligencia artificial generalmente esto no es admisible. La información simbólica tiene que estar presente de alguna manera.

En los problemas de inteligencia artificial es frecuente que se utilicen métodos de búsqueda de soluciones en un espacio de configuración más o menos grande, con objeto de encontrar un método óptimo (o al menos cuasi-óptimo) para obtener el objetivo deseado. A veces se realizan búsquedas exhaustivas, a lo ancho o en profundidad, que analizan todas las posibilidades y escogen, entre ellas, la mejor. Pero en general, si el proceso es complejo, el espacio de configuración es desmesurado y los métodos exhaustivos no son factibles, pues se producirá una *explosión combinatoria* en la que el número de caminos crece enormemente, desbordando la capacidad de cálculo de los ordenadores actuales, e incluso de todos los ordenadores posibles.

En la práctica, hay que utilizar métodos más rápidos que lleven en poco tiempo a una solución aceptable, aunque no sea precisamente la óptima. Estos métodos, que se apoyan generalmente en información procedente de la experiencia, se llaman por ello *heurísticos* e incorporan estrategias que limitan drásticamente la búsqueda, permitiendo encontrar una solución razonable para un problema, a pesar de que el espacio de búsqueda sea muy extenso.

De acuerdo con la definición 1, una aplicación de inteligencia artificial tiene que cumplir las condiciones siguientes:

- Que al menos parte de la información a tratar tenga carácter simbólico.
- Que el problema a resolver dé lugar a un espacio de búsqueda de soluciones (espacio de configuración) muy extenso, que no se trate de un problema trivial, que no sea posible resolverlo con una simple búsqueda exhaustiva.
- Que la forma más práctica de abordar el problema sea la utilización de reglas heurísticas basadas en la experiencia, que acorten el proceso de búsqueda de soluciones y eviten la explosión combinatoria.
- En principio, el programa debería ser capaz de extraer esas reglas heurísticas de su propia experiencia, es decir, debe ser capaz de aprender. Para ello se pueden utilizar técnicas como el reconocimiento de patrones (Bishop 2006), las cadenas de Markov (Norris 1998) o los algoritmos genéticos (Goldberg 1989), que en sí no se reducen a la inteligencia artificial, pues pueden aplicarse en muchos otros campos, pero pueden ser útiles también en este.

Es preciso distinguir la inteligencia artificial de otras tecnologías menos potentes, como las siguientes:

- *Inteligencia computacional*: es una versión *light* de la inteligencia artificial, que maneja datos esencialmente numéricos, de los que es capaz de extraer patrones utilizando conocimientos menos completos y exactos que los de las aplicaciones de inteligencia artificial (Siddique 2013).
- *Sistemas bio-inspirados*: son sistemas informáticos para la resolución de problemas complejos que se inspiran en conceptos biológicos, como la evolución (algoritmos genéticos) o el ADN (computación mediante ADN). No pueden considerarse inteligentes.

1.1 Las dos acepciones de la inteligencia artificial ↑

En lo que sí están de acuerdo todos los investigadores es en que los programas de inteligencia artificial actuales no permiten atribuir a las computadoras la capacidad de pensar y de actuar con verdadera inteligencia, en oposición a la apariencia de inteligencia, que obviamente es más fácil de conseguir. Por eso, el campo de la inteligencia artificial se divide en dos, uno que existe realmente, y el otro que de momento sólo es un objetivo inalcanzable, mejor o peor definido desde el punto de vista filosófico. Esos dos campos son los siguientes:

- *Inteligencia artificial débil*, que abarca todas las aplicaciones de las que disponemos hasta ahora, en las que la máquina actúa con apariencia de inteligencia, pero está claro que no piensa.
- *Inteligencia artificial fuerte*, que abarca a las hipotéticas máquinas programadas cuya inteligencia fuese comparable o superior a la humana.

2 Inteligencia artificial débil ↑

Todo lo que se ha logrado hasta ahora en este campo es claramente *inteligencia artificial débil*. Esto es lo que nos va a ocupar en esta parte del artículo. Últimamente se habla mucho de que la *inteligencia artificial fuerte* está a punto de conseguirse. En la tercera parte volveremos sobre este problema y analizaremos esas predicciones.

2.1 Breve historia de la inteligencia artificial ↑

Desde la más remota antigüedad, el hombre no ha dejado de buscar medios para disminuir el esfuerzo necesario para la realización de su trabajo. El primer paso en esta dirección se dio hace unos dos millones de años, con la invención de las armas y las herramientas. El segundo, hace al menos setecientos mil años, se plasmó en el dominio del fuego. Otro paso importante (la revolución neolítica) tuvo lugar hace unos diez mil años, con el comienzo de la agricultura y la ganadería. El ganado se utilizó desde el principio como un nuevo tipo de herramienta que permitía a su poseedor realizar más trabajo con menos esfuerzo. El cuarto paso trascendental fue la invención de la escritura, que tuvo lugar hace cosa de cinco mil años y abrió paso al arte literario y a la posibilidad de guardar información fuera de nuestro cerebro y de nuestro cuerpo, en papiro, pergamino, papel, etcétera.

Hace un poco más de doscientos años comenzó una nueva revolución tecnológica, la revolución industrial, cuyas posibilidades aún no se han agotado. En sus primeras fases, los siglos XVIII, XIX y la primera mitad del XX, esta revolución eliminó a los animales-herramienta que habían dominado la tecnología durante casi diez milenios, que fueron sustituidos por máquinas mecánicas propulsadas por fuentes de energía nuevas, como la térmica, la eléctrica, la química de los combustibles naturales y, ya en el siglo XX, la nuclear, cuya existencia ni se sospechaba a finales del siglo XIX.

Hacia la segunda mitad del siglo XX apareció un nuevo tipo de máquinas, las computadoras electrónicas, que ya no tratan de complementar el esfuerzo físico humano y extender el campo de acción de sus miembros. Su ámbito de aplicación es amplificar las actividades mentales del hombre. Estas máquinas realizan cálculos complejos a velocidades muy superiores a las nuestras, aunque suele decirse que son rígidas, que hay que preverlo todo y especificarlo claramente, porque en caso contrario nos encontraremos con resultados inesperados. Muy pocos programas exhiben un comportamiento que se pueda calificar de inteligente.

Casi desde el principio de la historia de la informática fue posible programar computadoras para actuar de una forma que sí suele considerarse inteligente. En 1956, Herbert Gelernter, del Laboratorio de IBM en Poughkeepsie, construyó un programa capaz de resolver teoremas de geometría plana, sorprendente para su época, que se considera uno de los primeros ejemplos de inteligencia artificial. Recuérdese que las computadoras pertenecían entonces a la primera generación y estaban construidas con válvulas de vacío.

Ante este ejemplo y otros parecidos que surgieron por entonces, los pioneros de la inteligencia artificial, encabezados por John McCarthy (1927-2011), se reunieron en un seminario en el Dartmouth College de Hanover (USA). Además de imponer nombre a la nueva disciplina (*inteligencia artificial*) lanzaron las campanas al vuelo y predijeron que en una década habría programas capaces de traducir perfectamente entre dos lenguas humanas y de jugar al ajedrez mejor que el campeón del mundo. Esto no sería más que el primer paso. Pronto sería posible construir máquinas capaces de comportarse con inteligencia igual o superior a la nuestra, con lo que entraríamos en una nueva vía en la evolución de nuestra sociedad. El viejo sueño de construir hombres artificiales se habría hecho realidad.

Pero las cosas no sucedieron como aquellos optimistas preveían. Es cierto que ese mismo año Arthur Samuel, de IBM, construyó un programa para jugar a las damas que guardaba información sobre el desarrollo de las partidas que jugaba y la utilizaba para modificar sus jugadas futuras (es decir, aprendía). En pocos años, tras un número suficiente de partidas, el programa fue capaz de vencer a su creador y desempeñaba un papel razonable en los campeonatos oficiales. Pero el juego de las damas es incomparablemente menos complicado que el ajedrez, y pronto se vio que este iba a ser un hueso bastante más duro de roer.

En los últimos años de la década de 1950, Alex Bernstein, de IBM, construyó un programa capaz de jugar al ajedrez como un principiante, que fue aireado por la prensa como el primer paso hacia el dominio del ajedrez por las computadoras. Pero el objetivo, ganar al campeón del mundo, o al menos desempeñar un buen papel en los torneos con jugadores humanos, se retrasaría más de 30 años respecto a la fecha prevista. En cuanto a la traducción de textos entre dos lenguas naturales, también resultó mucho más difícil de lo que se preveía, como se verá en el apartado siguiente.

El fracaso de las predicciones de los expertos provocó el desánimo de los investigadores en inteligencia artificial, muchos de los cuales se dedicaron a otras cosas. El problema se complicó en 1969, cuando Marvin Minski y Seymour Papert (Minski 1969) demostraron matemáticamente que las redes neuronales artificiales de una o dos capas (perceptrones sin capa oculta), con las que se venía investigando desde los años cincuenta, no pueden realizar una función Booleana tan sencilla como “o exclusivo” (la suma módulo 2). Esto sí se podía conseguir con una red sencilla de tres capas, pero los ordenadores de entonces no eran bastante rápidos ni tenían bastante memoria para trabajar con redes complejas de tres capas.

Durante los años setenta, el interés por la inteligencia artificial se renovó gracias a la aparición de los sistemas expertos. De nuevo se lanzaron las campanas al vuelo y se predijeron avances inmediatos demasiado ambiciosos. Arrastrado por esta tendencia, el gobierno del Japón puso en marcha a finales de los setenta el *proyecto de la quinta generación*, cuyo objetivo era ponerse en cabeza de la investigación informática mundial, desarrollando en diez años (siempre en diez años) máquinas capaces de *pensar* como los seres humanos, de comunicarse con nosotros en nuestra propia lengua, y de traducir perfectamente textos escritos en inglés y en japonés. Asustados por el proyecto, otros países lanzaron sus propios programas de investigación, aunque buscaron objetivos algo menos ambiciosos. A principios de los noventa, el proyecto japonés se dio por finalizado con un rotundo fracaso.

Aunque con altibajos, no siempre con resultados satisfactorios, los avances continuaron llegando en un goteo continuo. En 1997, 30 años después de lo previsto, se cumplió por fin el objetivo de que una máquina programada para jugar al ajedrez venciera al campeón del mundo. En el 2016, otra máquina venció al campeón mundial de Go, uno de los juegos más complejos que existen. También ha avanzado mucho la conducción automática de vehículos (coches y aviones), y se anuncia que los primeros coches sin conductor podrían estar en el mercado para el año 2020.

2.2 Aplicaciones de la inteligencia artificial débil ↑

Existen muchas aplicaciones, completamente diferentes unas de otras, que suelen clasificarse como *inteligencia artificial*. En algunas, los resultados han sido espectaculares y se aproximan a lo que entendemos intuitivamente por una máquina que piensa, aunque cuando se analizan a fondo se ve que la supuesta inteligencia no era tal, que se trata de aplicaciones programadas no muy diferentes de las que se utilizan en otros campos de la informática.

Son muchos los temas en los que ha sido posible aplicar técnicas de inteligencia artificial, hasta el punto de que este

campo se parece a un pequeño cajón de sastre. Veamos algunos:

- Algoritmos inteligentes para juegos
- Realización de razonamientos lógicos
- Reconocimiento automático de la palabra hablada
- Proceso de textos escritos
- Reconocimiento de imágenes y vehículos automáticos
- Sistemas expertos
- Redes neuronales artificiales
- Computación cognitiva y bases de conocimiento sobre el mundo

2.2.1 Algoritmos inteligentes para juegos ↑

En 1997, 30 años después de lo previsto, una máquina dedicada de IBM (Deep Blue) consiguió por fin vencer al campeón del mundo de ajedrez (Garry Kasparov) en un torneo a seis partidas (Campbell 2002). Sólo un año antes, la victoria de Kasparov contra el mismo programa había sido clara. En años sucesivos, los programas para jugar al ajedrez se han impuesto en torneos mixtos, como los dos *Campeonatos Mundiales del Hombre contra la Máquina* celebrados en Bilbao en 2004 y 2005 (Chess News 2004, 2005). En este caso, los tres programas se ejecutaron sobre máquinas comerciales y ganaron a los tres jugadores humanos de alta calificación por 8,5 a 3,5 en 2004, y por 8 a 4 en 2005.

Además del ajedrez, también se ha resuelto favorablemente la programación de otros juegos, como el backgammon o chaquete (Tesauro 1989), las damas (Schaeffer 2007), Jeopardy! (Ferrucci 2010), ciertas formas del póker (Bowling 2015) y el Go (BBC News Online 2016).

Sin embargo, los mismos autores de estos programas reconocen que, aunque sean capaces de ganar torneos, sus programas no son especialmente inteligentes. Ganan porque son más rápidos que los seres humanos y analizan un número inmenso de posibilidades, pero no utilizan la intuición como los campeones humanos, que saben dirigir sus esfuerzos hacia las líneas de ataque prometedoras, eludiendo las que no lo son. En este contexto, se ha llegado a decir que *Deep Blue* es capaz de ganar al ajedrez sin comprender el ajedrez (Hawkins 2004, 18).

2.2.2 Realización de razonamientos lógicos ↑

Existen tres formas principales de razonamiento humano:

- **Deducción**: es el método que más se aplica en las matemáticas. Proporciona una fiabilidad absoluta, pues si las premisas son correctas, la conclusión también tiene que serlo.
- **Inducción**: es el método que más se aplica en las ciencias de la naturaleza (la física, la química y algunas ramas de la biología). No proporciona fiabilidad absoluta, pero a medida que los resultados se confirman con experimentos independientes, su fiabilidad aumenta de forma considerable. Sin embargo, siempre queda un margen de inseguridad, pues queda abierta la posibilidad de que el próximo experimento que se realice contradiga la conclusión, por lo que se suele decir que en estas ciencias los descubrimientos son siempre provisionales (Popper 1962).
- **Abducción**: es el método que más se emplea en las ciencias humanas, la historia y algunas ramas de la biología, como la paleontología. Es el que proporciona menos fiabilidad, pues consiste en inferir la hipótesis

más sencilla que explique las observaciones. Para ello se establecen paralelos entre objetos, comportamientos o afirmaciones diferentes, y se busca documentación que los confirme, aunque por acumulación de indicios su fiabilidad puede llegar a ser grande (Alfonseca 2015).

Durante los años 60 y 70, el problema de programar ordenadores para que realicen deducciones lógicas se resolvió satisfactoriamente. Como consecuencia de estos estudios, se llegó a la conclusión de que el funcionamiento deductivo de la mente humana no es tan complicado como a primera vista parece, lo que tampoco fue una gran novedad, pues ya Aristóteles, en el siglo III antes de Cristo, había reducido los procesos de deducción a diecinueve formas elementales o silogismos. Primero se construyeron programas capaces de demostrar teoremas (Loveland 1978). Poco después, un programa que sólo tenía 165 instrucciones (Alfonseca 1975) pudo resolver correctamente los diecinueve silogismos clásicos. Para ello, y en síntesis, el programa convertía las proposiciones lógicas en relaciones de teoría de conjuntos y aplicaba reglas de reducción propias de este campo de las matemáticas.

En cambio, es mucho más difícil programar las computadoras para que realicen procesos de razonamiento inductivo o abductivo, por lo que estos campos de la investigación en inteligencia artificial continúan abiertos.

2.2.3 Reconocimiento automático de la palabra hablada ↑

En este campo se ha avanzado bastante desde los años setenta (Faundez 2000, Jurafsky 2008). Se trata de conseguir que los ordenadores entiendan la voz humana, para que sea posible darles órdenes de forma más natural, sin tener que utilizar una máquina de escribir eléctrica o un teletipo.

La investigación en este campo encontró dificultades en el hecho de que cada persona tiene su propia forma de pronunciar, ligeramente diferente de la de los demás, y en que el lenguaje hablado es más ambiguo que el escrito, pues a la existencia de palabras homófonas se suma el problema de que las palabras se funden unas con otras y se concatenan, resultando a veces difícil separarlas. Por esta razón, las primeras aplicaciones de proceso de voz ponían restricciones a las frases que son capaces de entender. Esas restricciones pertenecían a los tres grupos siguientes:

- Tamaño del diccionario: número de palabras que el programa es capaz de comprender.
- Separación de las palabras: algunos programas exigen que la persona que habla pronuncie las frases separando claramente las palabras entre sí.
- Existencia de una fase de educación del programa para adaptarse a la voz de una persona concreta, que será la única a la que normalmente podrá entender. Esta educación es un proceso laborioso, pues a menudo exige que la persona que va a utilizar el programa tenga que pronunciar al menos una vez todas las palabras del diccionario.

Las mejores aplicaciones sólo exigían una o dos de las restricciones anteriores. O bien se utilizan diccionarios de 2000 palabras, o se exige la separación de palabras, ampliando el diccionario hasta 20.000 términos. A menudo se alcanzaban grados de comprensión superiores al 90 por 100 de las palabras pronunciadas.

Actualmente se utilizan aplicaciones más robustas, sin las restricciones indicadas (Dong Yu 2015). Las empresas que se especializan en proceso de voz (Google, Microsoft, Nuance) aprovechan la enorme cantidad de datos existentes para entrenar sus aplicaciones con muchas personas diferentes. Google, por ejemplo, utiliza para ello el gran número de peticiones por voz que su buscador recibe cada día.

2.2.4 Proceso de textos escritos ↑

La investigación en proceso de textos se subdivide en dos áreas principales: proceso del lenguaje natural (Powers 1989, Manning 1999, Jurafsky 2008) y traducción automática (Hutchins 1992).

Nuestras lenguas adolecen de ambigüedad, tanto pragmática (cuando las dos personas que hablan no comparten el mismo contexto), semántica (una misma palabra puede tener varios significados, que suelen ser distintos en lenguas diferentes) como sintáctica (en una frase, la misma palabra puede desempeñar diversos papeles sintácticos).

Se ha hablado mucho de que los ordenadores del porvenir podrían programarse en lenguaje natural (castellano, inglés...). Actualmente esto es una utopía, aunque es posible dar al usuario la impresión de que se está comunicando en su lengua con el programa. Esto se consigue, generalmente, de tres maneras diferentes:

- Provocando la respuesta del usuario con preguntas prefabricadas y tratando de localizar palabras seleccionadas en dicha respuesta, sin hacer mucho caso de la sintaxis.
- Restringiendo el subconjunto del lenguaje natural que se puede utilizar, para eliminar ambigüedades.
- Restringiendo el tema de la conversación.

Un campo relativamente reciente es la *minería de datos*, cuyo objetivo es extraer información de textos escritos y tratar de comprender su significado (Manning 2008). Para ello se utilizan métodos estadísticos y se construyen corpus anotados que proporcionan información sobre los distintos términos, estableciendo relaciones entre unos términos y otros. Utilizando estos *corpora* los programas mejoran o aceleran la comprensión de los textos que deben interpretar.

Para averiguar a qué se refiere un texto mediante análisis automático se utilizan técnicas de *resolución de entidades*, también llamada *desambiguación de entidades*, con las que se alcanza actualmente entre un 80 y un 90% de identificaciones correctas. Un problema relacionado con este es la *correferencia*, que trata de deducir, en una frase pronominal, a qué sustantivo se refiere un pronombre.

En el campo de la traducción automática, los problemas se multiplican, pues en este caso los programas no tienen que enfrentarse con una sola lengua natural, sino con dos, ambas plagadas de ambigüedades e irregularidades, que además no coinciden casi nunca entre sí. Hacia finales de los años ochenta se propuso traducir primero las frases de la lengua de partida a un sistema de representación interna intermedio (Interlingua), que podría considerarse como una lengua artificial desprovista de ambigüedades. Después, otra parte del programa traductor trasladaría la traducción a Interlingua a una lengua diferente, utilizando las reglas y peculiaridades de ésta. Así se aislarían las dos lenguas, simplificando la traducción. Pero esta solución no ha llegado a implementarse. Actualmente Google está trabajando en la utilización de redes neuronales (véase la sección 2.2.7) para realizar traducciones directas entre dos lenguas naturales diferentes (Sutskever 2014), sin pasar por una Interlingua. A menos que se diga que el estado interno de la red neuronal es la Interlingua, aunque se trataría de una lengua *sui generis*, pues no tendría gramática reconocible.

Otra posibilidad, que se adoptó en el proyecto de traducción automática EUROTRA, patrocinado por la Unión Europea, y después en *Google Translate*, no tiene como objetivo realizar una traducción perfecta de los textos de partida, sino obtener una primera aproximación sobre la que un traductor humano puede trabajar para mejorarla, lo que le permite aumentar considerablemente su rendimiento (*traducción asistida por computadora*).

2.2.5 Reconocimiento de imágenes, y vehículos automáticos ↑

El reconocimiento de imágenes es otro de los campos en los que se aplican técnicas de inteligencia artificial. Cuando observamos una escena a través de la vista, somos capaces de interpretar la información que recibimos y separar la imagen en objetos independientes bien identificados. Este campo de investigación intenta programar máquinas y robots para que reconozcan visualmente los elementos con los que han de relacionarse.

Una de las aplicaciones más espectaculares de la visión de máquinas es el coche sin conductor. Este proyecto, bastante avanzado en la actualidad por parte de varias empresas, especialmente Google (Fisher 2013), tiene por objeto construir vehículos capaces de prescindir del conductor en el tráfico de las carreteras y las calles de una ciudad. Estas investigaciones, que comenzaron en la Universidad Carnegie Mellon durante los años ochenta (Jochem 1995), recibieron un fuerte impulso durante los noventa, cuando en 1995 un coche sin conductor diseñado por Ernst

Dieter Dickmanns recorrió 1758 km por las autopistas alemanas llevando un conductor humano para casos de emergencia, que tuvo que tomar el control un 5% del tiempo (Dickmanns 2007). En lo que llevamos del siglo XXI, la investigación en el campo del coche sin conductor ha seguido avanzando, y no parece lejano el momento en que se autorice su comercialización.

Algunas otras aplicaciones avanzadas de proceso de imágenes:

- *Google Photos*, una aplicación que permite hacer búsquedas en Internet basándose en imágenes, en lugar de textos.
- OCR (*Optical Character Recognition*), extracción de textos a partir de fotografías. Combinado con *Google Translate* (véase la sección 2.2.4), permite además traducirlos a otros idiomas.
- *Realidad aumentada*, campo que aún no ha alcanzado todos sus objetivos, que intenta suplementar las percepciones de los sentidos humanos con información generada por un ordenador (imágenes, sonidos, o datos GPS). Esto se conseguiría mediante cascos, lentes de contacto o dispositivos parecidos a gafas.

2.2.6 Sistemas expertos ↑

Una vez que se comprobó que no era tan difícil conseguir que los programas realizaran deducciones lógicas, el camino estaba abierto hacia una aplicación práctica de la inteligencia artificial: los sistemas expertos (Nilsson 1998, cap. 17.4). El primer intento en esta dirección lo realizaron hacia 1965 Edward A. Feigenbaum y Joshua Lederberg en la Universidad norteamericana de Stanford (Feigenbaum 1983). Se trataba de construir un programa capaz de realizar deducciones inteligentes a partir de datos de análisis químico (espectrogramas de masas) para obtener la fórmula desarrollada de compuestos orgánicos desconocidos. Después de varios años de trabajo, los investigadores terminaron con éxito el proyecto. Su programa, llamado DENDRAL, se utilizó durante décadas en universidades y laboratorios de análisis de todo el mundo.

Durante los años setenta y ochenta, la investigación en sistemas expertos se aplicó en campos muy variados: diagnóstico médico, matemáticas, física, prospecciones mineras, genética, fabricación automática, configuración automática de computadoras...

¿Qué es un sistema experto y en qué se diferencia de los programas ordinarios? Mientras en estos el conocimiento se organiza en dos niveles, las instrucciones y los datos, en los sistemas expertos hay tres componentes:

- Un sistema de inferencia capaz de hacer deducciones lógicas sobre los datos de que dispone.
- Un conjunto de reglas, utilizadas por el sistema de inferencia para realizar deducciones o acciones determinadas.
- Los datos sobre el problema concreto que se quiere resolver.

Uno de los primeros sistemas expertos fue MYCIN, que diagnosticaba enfermedades infecciosas del aparato circulatorio y recomendaba un tratamiento. Los datos del problema eran los síntomas del paciente, los resultados de sus análisis, y otras cuestiones clínicas que tuviesen que ver con su caso concreto. Las reglas de deducción que se utilizaban para realizar el diagnóstico y proponer el tratamiento se le proporcionaron a MYCIN durante su construcción, no las aprendía, siempre utilizaba las mismas reglas en todos los problemas que tenía que resolver. Por último, el sistema de inferencia era un programa capaz de aplicar las reglas a los datos para obtener conclusiones razonables.

El conjunto de reglas de deducción de un sistema experto se llama su *base de conocimientos*. Para representarlas suelen utilizarse *reglas de producción*: expresiones de la forma **SI condición ENTONCES acción**. La especificación de las reglas evitando toda ambigüedad es el problema más importante con el que se enfrenta quien desee construir un sistema experto.

El sistema de inferencia realiza búsquedas en el espacio de conocimientos y de datos. A menudo es imposible probar

todas las posibilidades, pues su número es excesivo, por explosión combinatoria. Por eso existen diversos tipos de búsqueda: en *anchura*, en *profundidad*, o *heurística*. La última hace uso de información *ad-hoc* para decidir el camino a seguir. La búsqueda puede ser también *hacia adelante*, partiendo del estado inicial y aplicando las reglas de deducción hasta llegar al objetivo, o *hacia atrás*, partiendo del estado final y remontándose hasta los conocimientos existentes, aplicando a la inversa la información de la base de conocimientos.

En un sistema experto resulta conveniente disponer de la posibilidad de hacer que explique sus deducciones. Cuando el sistema responde a una pregunta sobre un diagnóstico médico, por ejemplo, debe ser capaz de señalar cómo llegó a esa conclusión (qué reglas ha aplicado).

A partir de finales de los años ochenta, coincidiendo con el fracaso de la quinta generación japonesa, los sistemas expertos entraron en decadencia. Aunque no han desaparecido, hoy no desempeñan el papel principal en la investigación en inteligencia artificial.

2.2.7 Redes neuronales artificiales ↑

Esta es una de las aplicaciones más antiguas de la inteligencia artificial (Gurney 1997). También ha sido una de las más sometidas a exageraciones y previsiones insólitas. Son redes inspiradas en los sistemas nerviosos de los animales, formadas por muchas componentes interconectadas capaces de cierta actividad computacional. Las *neuronas* que componen estas redes están generalmente bastante simplificadas, en comparación con las que forman parte del sistema nervioso humano y de muchos animales. Se ha dicho que estas redes son capaces de resolver los problemas más difíciles que existen, en principio (los problemas NP-completos del tipo del viajante de comercio y otros equivalentes), que un ordenador normal sólo puede resolver en un tiempo que crece exponencialmente en función de la complicación del problema. Y hasta cierto punto es verdad, siempre que tengamos en cuenta que la solución obtenida no es necesariamente la óptima, sino tan sólo una aproximación, que muchas veces es suficiente para nuestras necesidades.

Las primeras redes neuronales fueron definidas por Warren McCullogh y Walter Pitts (McCulloch 1943). En la década siguiente, Frank Rosenblatt ideó el perceptrón (Rosenblatt 1958), una red neuronal de dos capas (capa de entrada y capa de salida), capaz de aprender qué respuesta debe corresponder a una entrada concreta. La investigación en este campo se estancó tras la publicación del artículo de Minski y Papert antes mencionado (Minski 1969), que demostró que un perceptrón de dos capas no es capaz de resolver la función *o-exclusivo*. Algunos años más tarde, con la introducción de una tercera capa de neuronas en la red neuronal (situada entre las capas de entrada y de salida) y con la invención del algoritmo de propagación hacia atrás, se resolvió el problema de la función *o-exclusivo* y la investigación en el campo de las redes neuronales volvió a avanzar.

En la actualidad (Kruse 2013) se está trabajando en varias direcciones: a) La implementación de las redes neuronales mediante dispositivos de hardware que hacen uso de la nanotecnología, en lugar de simularlas en un ordenador. b) Nuevos algoritmos para resolver, mediante redes neuronales, problemas de *reconocimiento de patrones* y aprendizaje automático. c) Implementación de redes basadas en neuronas más complejas, inspiradas en las neuronas biológicas.

2.2.8 Computación cognitiva y bases de conocimiento sobre el mundo ↑

Uno de los problemas que han dificultado la investigación en inteligencia artificial ha sido el hecho de que las computadoras apenas poseen conocimientos sobre el mundo que nos rodea, lo que les pone en desventaja evidente respecto a cualquier ser humano, que sí posee esa información, pues la ha adquirido desde su infancia y puede utilizarla para resolver problemas de sentido común que a nosotros nos parecen triviales, pero que son dificilísimos de resolver para las máquinas que no disponen de la información necesaria.

Un primer paso en esa dirección podría ser Wolfram|Alpha (Wolfram 2012). Mientras los buscadores de Internet (como Google) reaccionan ante una pregunta proporcionando una serie de direcciones de Internet donde el lector puede

(quizá) encontrar la respuesta a su pregunta, Wolfram|Alpha intenta proporcionar directamente la respuesta, extrayéndola de la información disponible en la web y de la que ha ido acumulando en su *base de conocimientos* a lo largo del tiempo.

También en esta línea, IBM ha puesto en marcha un proyecto de computación cognitiva llamado DeepQA (IBM Research 2011), cuyo objetivo es construir una computadora (Watson) que, a partir de datos muy generales y abundantes (***big data***) y utilizando técnicas de inteligencia artificial y aprendizaje automático, sea capaz de hacer predicciones e inferencias útiles, y de responder a preguntas expresadas en lenguaje natural.

Por el momento, estos sistemas no pueden ser tan generales como los seres humanos, y normalmente se restringen a uno o unos pocos campos de aplicación concretos.

3 Inteligencia artificial fuerte ↑

El pasado de la inteligencia artificial ha estado sometido a numerosos altibajos, cuya causa quizá deba buscarse en el nombre mismo de la disciplina: ***Inteligencia artificial*** sugiere con facilidad las ideas de ***hombre artificial***, o ***el hombre igualado o superado por la máquina***, que pertenecen más bien a la controversia filosófica que a la tecnología informática. Estos términos están cargados de contenido emotivo y despiertan fuertes rechazos o adhesiones casi fanáticas. Desde mediados del siglo XX vemos continuamente que los medios de comunicación y algunos investigadores en informática (no todos, desde luego) lanzan las campanas al vuelo y anuncian que la *inteligencia artificial fuerte*, la construcción de máquinas programadas tan inteligentes o más que nosotros, está a punto de conseguirse. Sin duda, todo avance nuevo en inteligencia artificial es útil y valioso, pero se arriesga a verse despreciado, porque el listón está demasiado alto.

Para plantearse correctamente el problema, lo primero que habría que hacer es definir lo que se entiende por *inteligencia humana*. Y ahí precisamente nos encontramos con las primeras dificultades. Sabemos que nuestra inteligencia está ligada de algún modo con el funcionamiento de nuestro cerebro, pero no sabemos cómo funciona nuestro cerebro. En palabras de Jeff Hawkins:

Antes de tratar de construir máquinas inteligentes, tenemos que comprender primero cómo piensa el cerebro, y en eso no hay nada artificial. (Hawkins 2004, 4-5).

En un campo tan sujeto a controversia, no nos puede extrañar que ni siquiera una cuestión tan primordial como la anterior sea aceptada por todos los investigadores. De hecho, hay muchos que opinan que el problema de la *inteligencia artificial fuerte* no tiene nada que ver con la biología y es puramente tecnológico. Pero, como dice Hawkins:

Durante décadas, los científicos del campo de la inteligencia artificial han sostenido que los computadores serán inteligentes cuando alcancen una potencia suficiente. Yo no lo creo, y explicaré por qué: los cerebros y las computadoras hacen cosas fundamentalmente diferentes. (Hawkins 2004, 5).

A pesar de todo, Hawkins es optimista y piensa que las máquinas tan inteligentes como el hombre están a la vuelta de la esquina:

¿Podremos construir máquinas inteligentes?... Sí. Podremos y lo haremos. En las próximas décadas veremos una rápida evolución en las capacidades de esas máquinas en direcciones interesantes. (Hawkins 2004, 7).

Entre los que hacen predicciones optimistas destaca Ray Kurzweil, que lleva décadas anunciando la inminencia de la inteligencia artificial fuerte:

- En su libro *La era de las máquinas inteligentes* (Kurzweil 1990) parte de la base de que un programa de ordenador suficientemente avanzado exhibiría automáticamente inteligencia similar a la humana (justo lo que Hawkins niega en la segunda cita anterior). Algunas de sus predicciones en este libro ya han sido falsadas,

como el *teléfono traductor* para la primera década del siglo XXI.

- En el libro *La era de las máquinas espirituales* (Kurzweil 1999) predice que las máquinas más inteligentes que los hombres serán el resultado automático de la evolución de los computadores actuales durante un par de décadas. Sólo faltan cuatro años para que esta predicción quede asimismo falsada.
- Seis años después, Kurzweil publicó otro libro, *La singularidad está cerca* (Kurzweil 2005), en el que afirma que la confluencia de los avances de diversos campos (informática, robótica, genética y nanotecnología darán lugar para 2045 a una singularidad tecnológica, un avance tan enorme que ya no seremos capaces de comprenderlo, lo que equivale a decir que seremos superados por nuestras máquinas para siempre.
- Siete años después, Kurzweil publicó *Cómo crear una mente: revelado el secreto del pensamiento humano* (Kurzweil 2012), en el que sostiene que una inteligencia artificial mayor que la humana podría crearse en breve (a fines de la década de 2020), simplemente profundizando algunas técnicas típicas de la inteligencia artificial débil, como los modelos de Markov y los algoritmos genéticos.

Kurzweil no sólo parece obsesionado por la inteligencia artificial fuerte, también lo está por el problema de la inmortalidad humana, que ve también inminente (alguna vez la ha predicho para 2035), porque cree que en el fondo ambos problemas están estrechamente relacionados. Según él, la inmortalidad se alcanzará por la confluencia de tres caminos diferentes:

1. Gracias a los avances de la medicina. Cuando los investigadores sean capaces de aumentar nuestra esperanza de vida un año cada año, seremos automáticamente inmortales. Desgraciadamente para Kurzweil, las predicciones de la ONU para el siglo XXI (United Nations 2015) apuntan a una disminución progresiva del incremento de la esperanza de vida, en lugar de a un aumento.
2. Mediante la combinación de los avances de la medicina y la informática, a través de la construcción de órganos artificiales. El hombre del futuro se transformaría en *cyborg* y sería prácticamente inmortal. Esta es la línea seguida por los *transhumanistas* (Bostrom 2005). El problema aquí está en si será posible sustituir el cerebro humano por un cerebro artificial. Si no lo es, la supuesta inmortalidad terminaría en cuanto el cerebro se deteriorase. Y si lo fuese, ¿acaso quien se sometiese a esa sustitución seguiría siendo el mismo ser humano?
3. A través de la inteligencia artificial fuerte. Cuando seamos capaces de construir superinteligencias, podremos descargar nuestra consciencia en una de ellas, para seguir viviendo indefinidamente. Sin embargo, es preciso reconocer que este objetivo parece demasiado ambicioso, al menos por ahora. Si no sabemos qué es la consciencia, ¿cómo vamos a descargarla?

Otros investigadores no son tan optimistas como Hawkins y Kurzweil. Así, Ramón López de Mántaras dice:

La otra vertiente, la de inteligencia artificial generalista, intenta desarrollar inteligencias artificiales que tengan [la] versatilidad y [la] capacidad general de saber de muchas cosas. Esto no significa, y ahí es donde está el error de algunos de estos planteamientos, que esa inteligencia tenga que ser igual que la humana. De hecho, es imposible, en mi opinión. Por muy sofisticadas que sean algunas inteligencias artificiales en el futuro, dentro de 100.000 o 200.000 años, serán distintas de las humanas. (López de Mántaras 2013).

Como vemos, las discrepancias son impresionantes: mientras unos hablan de unas pocas décadas, otros lo dejan para dentro de algunos cientos de miles de años. Como veremos en la sección 3.5, además de estas dos posturas (a muy corto y muy largo plazo) existe una tercera: la que afirma, por razones filosóficas, que el objetivo de construir máquinas más inteligentes que el hombre es probablemente imposible.

Puesto que los criterios puramente tecnológicos son anteriores a los filosóficos, vamos a ver primero con detalle el más antiguo de todos, la *prueba de Turing*.

3.1 La prueba de Turing ↑

En 1950, adelantándose a su época, el matemático y químico inglés Alan Turing intentó definir las condiciones en que

sería posible afirmar que una máquina es capaz de pensar como nosotros. Para Turing, esto se conseguirá cuando la máquina sea capaz de engañar a los seres humanos, haciéndoles pensar que es uno de ellos (Turing 1950). Esta prueba se llama *el juego de la imitación*.

Para desarrollar su teoría, lo primero que hizo Turing fue proponer una prueba preliminar. Se trata de que un investigador se comunique con dos personas de distinto sexo y descubra cuál de ellas es el hombre y cuál es la mujer, haciéndoles preguntas y estudiando las respuestas que recibe. Los dos sujetos se encuentran, naturalmente, fuera de su vista, y se comunican con él a través de teletipos o terminales de ordenador. Uno de ellos (el hombre) trata de engañarle, el otro (la mujer) trata de ayudarle. ¿Cuántas preguntas tendrá que realizar para descubrir quién es quién? ¿En qué porcentaje de casos conseguirá engañarle el hombre? Turing estimó que dicho porcentaje no rebasaría mucho el 30%, o sea, que en un 70% de los casos el investigador no se dejaría engañar. Es curioso que estas cifras se hayan dado por buenas sin comprobación alguna. Al menos, yo no soy consciente de que este experimento se haya llevado a la práctica.

En la segunda parte de la prueba, se sustituye el hombre por una computadora convenientemente programada. El segundo participante puede ser un ser humano cualquiera. Se trata de que el investigador descubra cuál de sus dos contertulios es la computadora. ¿Varían las circunstancias respecto al problema anterior? ¿Es capaz el interrogador de descubrir cuál es la máquina, haciendo menos preguntas que en el caso del hombre y la mujer?

Según Turing, se podrá afirmar que una máquina piensa cuando los resultados de las dos pruebas sean idénticos. Si una máquina que intenta hacerse pasar por humana fuese capaz de engañar a los seres humanos con la misma facilidad con que un ser humano puede engañar a otro, habría que considerarla inteligente y equivalente a los seres humanos.

Turing no se limitó a plantear la prueba, sino que hizo predicciones concretas:

Yo creo que en unos cincuenta años será posible programar computadoras, con una capacidad de almacenamiento de alrededor de 10^9, para que sean capaces de jugar tan bien al juego de la imitación que un interrogador promedio no tendrá más del 70 por ciento de probabilidad de hacer la identificación correcta después de cinco minutos de interrogatorio. (Turing 1950).

Durante muchos años, ningún programa se acercó siquiera a resolver la prueba de Turing. Curiosamente, el que más se aproximó a ello fue ELIZA (Weizenbaum 1966), que se hacía pasar por un psiquiatra que dialoga con sus supuestos pacientes. Pero sólo los *pacientes* más inocentes se dejaban engañar por él, bastaba con intercambiar media docena de frases para descubrir que estabas hablando con una computadora, por la forma rígida en que contestaba.

En una prueba realizada en 2014, la predicción de Turing pareció cumplirse con 14 años de retraso, cuando un *chatbot* (un programa que toma parte en una conversación de *chat*) llamado *Eugene Goostman* consiguió convencer al 33% de sus contertulios, tras cinco minutos de conversación, de que era un chico ucraniano de 13 años. Sin embargo, algunos analistas no ven las cosas tan claras. El hecho de que el programa se hiciese pasar por un adolescente extranjero, en lugar de un compatriota adulto, aumentó el nivel de credulidad de sus contertulios en el *chat*. Comentando este resultado, Evan Ackerman escribió:

La prueba de Turing no demuestra que un programa sea capaz de pensar. Más bien indica si un programa puede engañar a un ser humano. Y los seres humanos somos realmente tontos. (Ackerman 2014).

Muchos investigadores piensan que la prueba de Turing no basta para definir o detectar la inteligencia. Por una parte, intenta demostrar que hay inteligencia, sin definirla. Por otra, la prueba se apoya en decisiones tomadas por personas concretas, cuyo juicio puede no ser de confianza, como señala Ackerman. Actualmente, ni filósofos ni informáticos consideran que la prueba de Turing tenga verdadero valor, por lo que los intentos de llevarla a cabo (como el mencionado) deberían considerarse simplemente anecdóticos.

Algunos investigadores (Legg 2007, Hernández-Orallo 2010) han propuesto pruebas alternativas a la de Turing para detectar la posible inteligencia de las máquinas, pero por el momento no se ha impuesto ninguna.

3.2 La habitación china ↑

En 1980, el filósofo John Searle propuso una nueva prueba, *la habitación china* (Searle 1980, 1999). Veamos en qué consiste:

1. Supongamos que tenemos un programa de computadora capaz de pasar satisfactoriamente la prueba de Turing, que se pone a dialogar con una persona china. En la conversación, los dos participantes utilizan caracteres chinos para comunicarse por escrito. La computadora, que está encerrada en una habitación para que la persona no la vea, lo hace tan bien que es capaz de engañarla, por lo que la persona cree estar dialogando con un ser humano que conoce perfectamente la lengua china.
2. Ahora Searle saca de la habitación la computadora y en su lugar se coloca él mismo. Searle reconoce que no sabe chino, pero se lleva un organigrama del programa que utilizó la computadora para dialogar con la otra persona. En principio, utilizando ese programa, Searle sería capaz de dialogar con ella en su propia lengua tan bien como lo hacía la computadora. Cada vez que reciba un texto escrito en chino, aplica las reglas y escribe los signos correspondientes a la respuesta que habría dado la computadora.
3. Searle sabe que no sabe chino. Sabe también que no se ha enterado de la conversación que acaba de mantener con la otra persona, a pesar de que dicha conversación fue capaz de engañarla, haciéndola creer que ha estado dialogando con un ser humano que sabe chino.
4. Como la actuación de la computadora fue idéntica a la de Searle, es de suponer que la máquina tampoco entendió la conversación. Ahora Searle plantea la siguiente pregunta: ¿Es consciente de ello la computadora, como Searle lo es?

Es de suponer que la computadora no es consciente de ello, como ningún programa actual lo sería. Luego no basta con que una máquina sea capaz de pasar la prueba de Turing para que se pueda afirmar que piensa, para poder considerarla inteligente. Hacen falta dos cosas más: que comprenda lo que escribe y que sea consciente de la situación. Mientras eso no ocurra, no podremos hablar estrictamente de *inteligencia artificial fuerte*.

En todo esto subyace un problema muy importante: para construir una inteligencia artificial fuerte, parece necesario dotar a las máquinas de consciencia. Pero si no sabemos qué es la consciencia, ni siquiera la nuestra, ¿cómo vamos a conseguirlo? En los últimos tiempos se han realizado muchos avances en neurociencia, pero aún estamos muy lejos de poder definir lo que es la consciencia y saber de dónde surge y cómo funciona, así que mucho menos podemos crearla, ni siquiera simularla.

El argumento de la habitación china no ha convencido a todo el mundo. Las respuestas que se le han planteado son muy variadas y pueden clasificarse en varios grupos, de los que aquí sólo se mencionan tres:

- Los que sostienen (Dennett 1991, Russell 2003) que en esta prueba sí hay alguien (o algo) que entiende chino: la habitación completa, el sistema, el conjunto, aunque el ser humano que forma parte del sistema no lo entienda. Searle replica que su argumento puede simplificarse, porque él podría -en principio- aprender de memoria el organigrama del programa, con lo que el sistema completo se reduciría a él mismo, que sigue sin entender chino y lo sabe. De este contra-argumento hay varias versiones, más o menos equivalentes.
- Otros, como Marvin Minski (1980) sostienen que la mera presencia del software hace aparecer una *mente virtual*, diferente del sistema, y esta mente sí entiende chino. Searle responde que dicha mente no es tal, sino una simple simulación.
- Algunos (ver Cole 2004) aducen que el programa que permite a la computadora contestar a la persona china tiene que contener una gran cantidad de información sobre el mundo en general, y que es esto lo que da sentido a los símbolos. Searle contesta que no cree que sea posible introducir esa información contextual en un programa.

3.3 El problema de la contención ↑

Si la inteligencia artificial fuerte fuese practicable, se nos plantearía un problema importante:

Definición 2: *El* **problema de la contención** *puede definirse con esta pregunta: ¿Es posible programar una superinteligencia de tal manera que no le esté permitido causar daño a ningún ser humano?*

En 1942, en el contexto de sus cuentos cortos sobre robots, Isaac Asimov formuló las tres leyes de la robótica (Asimov 1950):

1. Un robot no debe hacer daño a un ser humano o, por inacción, permitir que un ser humano sufra daño.
2. Un robot debe obedecer las órdenes dadas por los seres humanos, excepto si estas órdenes entrasen en conflicto con la primera ley.
3. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o la segunda ley.

Esencialmente, la primera ley de Asimov es equivalente al problema de la contención. Pues bien, hay indicios matemáticos de que no es posible resolver el problema de la contención. En particular, en un estudio muy reciente (Alfonseca inédito) se demuestra que el problema de la contención es equivalente al problema de la parada, que Alan Turing demostró que no tiene solución (Turing 1937). Si esto se confirma, tenemos dos posibilidades:

a) Renunciar a crear superinteligencias, para evitar el posible mal que podrían causarnos.

b)Renunciar a estar seguros de que las superinteligencias no podrán causarnos daño. Si fuese posible llegar hasta ese punto en la investigación en inteligencia artificial, y si decidimos hacerlo, habrá que asumir un riesgo.

Norbert Wiener, fundador de la Cibernética (una tecnología interdisciplinar para la exploración y el diseño de sistemas auto-regulados), compara con la magia el comportamiento literal y poco inteligente de las computadoras, y pone el siguiente ejemplo (Wiener 1948-1961):

Más terrible que cualquiera de estos cuentos es la fábula de la pata del mono, escrita por W.W.Jacobs, escritor inglés de principios del siglo [XX]. Un trabajador inglés retirado está sentado a la mesa con su esposa y un amigo, un sargento británico que acaba de volver de la India. El sargento muestra a sus anfitriones un amuleto en forma de pata de mono seca y marchita... [que tiene] el poder de conceder tres deseos a tres personas... El último [deseo de su primer propietario fue] morir... Su amigo... desea poner a prueba sus poderes. Como primer [deseo] pide 200 libras. Poco después llaman a la puerta y un funcionario de la empresa en que trabaja su hijo entra en la habitación. El padre se entera de que su hijo ha muerto en un accidente con las máquinas, por lo que la empresa... desea pagarle al padre la suma de 200 libras... Desconsolado, el padre formula su segundo deseo -que su hijo vuelva- y cuando se oye otra llamada en la puerta... aparece algo... el fantasma del hijo. El último deseo es que el fantasma desaparezca. La moraleja de estas historias es que la magia es literal... Las máquinas que aprenden también son literales. Si programamos una máquina... y le pedimos que nos lleve a la victoria, y no sabemos lo que queremos decir con eso, veremos al fantasma llamando a nuestra puerta.

La moraleja que Wiener extrae de este cuento (Jacobs 1902) es evidente: como la magia, las computadoras obedecen literalmente las órdenes que reciben, sin tener en cuenta muchos datos de los que nosotros disponemos, que solemos dar por supuestos, pero que ellas no poseen.

3.4 Algunos proyectos fallidos hacia la inteligencia artificial fuerte ↑

Es curioso que los proyectos más ambiciosos y famosos del campo de la inteligencia artificial, los que se han adentrado en las zonas fronterizas entre la inteligencia artificial débil y la fuerte, hayan terminado mal desde el punto de vista de la tecnología, aunque no desde otros puntos de vista. Vamos a revisar someramente tres de los más conocidos:

3.4.1 Quinta generación japonesa ↑

Hacia los años setenta se distinguían cuatro generaciones de computadoras, tanto desde el punto de vista del *hardware* como del *software*.

Generaciones de *hardware*:

Primera generación: ordenadores construidos con válvulas electrónicas de vacío.

Segunda generación: ordenadores construidos con transistores.

Tercera generación: ordenadores construidos con circuitos integrados.

Cuarta generación: ordenadores dotados de un sistema operativo capaz de funcionar en tiempo compartido (ejecutando varias aplicaciones al mismo tiempo) y con memoria virtual.

Generaciones de *software*:

Primera generación: programas escritos en lenguaje binario (lenguaje de la máquina).

Segunda generación: programas escritos en lenguaje simbólico (ensamblador).

Tercera generación: programas escritos en lenguajes de alto nivel. El primero fue FORTRAN, pero luego proliferaron como en una torre de Babel.

Cuarta generación: programas escritos mediante sistemas de generación automática de aplicaciones, con acceso a bases de datos.

A finales de los años setenta, el gobierno japonés decidió emprender un proyecto que pusiera a su país a la cabeza de la informática mundial. Según explicaron, estaban cansados de que el resto del mundo considerara a los japoneses como *copiadores eficientes* de la tecnología desarrollada por otros países. Esta vez querían ser ellos los copiados. Por eso pusieron en marcha el proyecto de la **quinta generación** (tanto en el *hardware* como en el *software*), que consistiría en lo siguiente:

Quinta generación de *hardware*: ordenadores adaptados para simplificar la construcción de aplicaciones de inteligencia artificial.

Quinta generación de *software*: programas de inteligencia artificial escritos en lenguaje Prolog, capaces de interaccionar con el usuario en su propia lengua (inglés y japonés) y de traducir correctamente entre esas dos lenguas.

El lenguaje Prolog, inspirado en las reglas de los sistemas expertos descritas en el apartado 2.6, fue diseñado durante los años setenta para escribir aplicaciones de inteligencia artificial (Clocksin 1981-2003), y se basaba en la realización de deducciones lógicas representadas por reglas definidas en orden arbitrario, por lo que se dice que este lenguaje es *no procedimental*. Un gestor de inferencias como el de los sistemas expertos se encarga de decidir en qué orden deben ejecutarse las reglas.

El proyecto de la quinta generación debía durar diez años y terminó a principios de los años noventa con un fracaso. Los supuestos ordenadores de quinta generación que iban a construirse resultaron ser ordenadores personales corrientes, dotados de un *firmware* que les permitía entender el lenguaje Prolog, lo cual no era nuevo, pues los primeros ordenadores personales llevaban un *firmware* que les capacitaba para entender el lenguaje Basic. Los grandes objetivos (traducción automática y comprensión del lenguaje natural) no fueron alcanzados. El éxito del proyecto consistió en empujar a otros países a lanzar proyectos menos ambiciosos, algunos de los cuales sí dieron lugar a resultados razonables.

3.4.2 Iniciativa de Defensa Estratégica ↑

Conocida también como **SDI** (por sus siglas en inglés) y como **Star Wars** (*la guerra de las galaxias*), este proyecto fue lanzado por la administración de Ronald Reagan durante los años 80 para romper el equilibrio nuclear entre las grandes potencias. Su objeto era la construcción de un sistema de defensa espacial que protegiera a los Estados Unidos contra un posible ataque global con proyectiles nucleares, gracias a un conjunto de satélites artificiales que escudriñarían permanentemente el planeta para localizar los primeros síntomas de un ataque nuclear y pondrían en marcha diversos mecanismos dirigidos a contrarrestar dicho ataque, impidiendo que los proyectiles intercontinentales ICBM (*Inter-Continental Balistic Missile*) llegaran a su destino. Un ICBM tiene un alcance de miles de kilómetros y sería lanzado en una órbita que le llevaría fuera de la atmósfera, hasta una altitud de unos 1200 kilómetros. De acuerdo con los tratados contra la proliferación nuclear, cada proyectil no podría transportar más de diez cabezas nucleares, que al final se separarían para dirigirse hacia diez objetivos diferentes, aunque sí podría llevar un número mucho más grande de señuelos, proyectiles secundarios sin carga nuclear, cuya única misión era disminuir la probabilidad de que los proyectiles cargados fuesen destruidos antes de alcanzar el objetivo.

Un ataque nuclear total supondría el lanzamiento de unas 10.000 cabezas nucleares. Antes de la puesta en marcha de la iniciativa de defensa estratégica, cuando una de las grandes potencias descubriera que la otra había lanzado el ataque, sólo le quedaban dos posibilidades: aceptar la destrucción sin tomar represalias o lanzar un contragolpe, lo que daría lugar a la destrucción mutua de ambas partes. El mundo permaneció en esta situación de amenaza permanente durante unos cuarenta años.

El objetivo de la iniciativa de defensa estratégica era proporcionar una tercera alternativa: localizar prematuramente el ataque y destruir las 10.000 cabezas nucleares durante su vuelo por el espacio. Para conseguirlo sería preciso detectar los ICBM durante el lanzamiento (los diez primeros minutos) y destruirlos durante la fase de movimiento balístico sin empuje activo (unos 25 minutos), pero antes de la fase terminal, cuando las cabezas nucleares y los señuelos se independizan, lo que haría prácticamente imposible la destrucción de todos ellos.

Para poder destruir los proyectiles durante la fase de movimiento balístico, se iban a utilizar armas avanzadas (algunas aún no has sido desarrolladas), como proyectiles antiproyectiles, armas de haces de partículas o láseres de rayos X. Obviamente no hay tiempo para la participación humana en el proceso, por lo que el sistema entero debería ser autónomo, controlado mediante programas de *inteligencia artificial*.

El proyecto SDI (*Star Wars*) no llegó a implementarse, pero el simple anuncio de su puesta en marcha quizá fue uno de los elementos que influyó para que se rompiera el equilibrio entre las dos grandes potencias, lo que habría conducido a la desintegración de la Unión Soviética y a la liberación de los países pertenecientes al Pacto de Varsovia, que se deshicieron de los regímenes dictatoriales comunistas y pasaron al sistema democrático, realineándose después como miembros de la Unión Europea y de la OTAN.

3.4.3 Cápsulas inteligentes para la exploración del sistema solar ↑

La Agencia Espacial estadounidense (NASA) ha utilizado también técnicas de inteligencia artificial para la exploración del espacio. En la exploración de astros lejanos las comunicaciones son difíciles, pues el tiempo de transmisión puede ser superior a una hora entre ida y vuelta, por lo que, o bien se programa todo por anticipado, o es preciso introducir técnicas que permitan a la cápsula espacial responder automáticamente a situaciones no previstas.

El diseño de la primera cápsula espacial que debería ser completamente autónoma comenzó en 1982. Su objetivo era explorar Titán, el más grande de los satélites de Saturno. El grupo de trabajo que inició el diseño escribió esto:

Los sistemas de inteligencia artificial con capacidad de formación automática de hipótesis serán necesarios para el examen autónomo de ambientes desconocidos. Esta capacidad es muy deseable para la exploración eficaz del Sistema Solar, y es esencial para la investigación de otros sistemas estelares.

Las técnicas de inteligencia artificial que se consideraban necesarias eran:

1. Corrección autónoma de errores.
2. Proceso en paralelo.
3. Capacidad lógica y dialéctica.
4. Adquisición, reconocimiento de formas y formación de conceptos.
5. Utilización del razonamiento abductivo.

Cuando en 1997 se lanzó la misión Cassini/Huygens (NASA 2011), se programó por procedimientos tradicionales, pues varias de estas técnicas aún no estaban disponibles (siguen sin estarlo). Cuando la misión llegó en 2004 a las proximidades de Saturno, una de sus componentes, la sonda Huygens, descendió a la superficie de Titán y envió numerosos datos sobre lo que encontró allí. Es curioso, sin embargo, que la actuación de la cápsula tuvo que modificarse en el último momento desde la Tierra, pues se descubrió que el programa que la dirigía contenía un error crítico de diseño (no se había tenido en cuenta el efecto Doppler). Como era imposible cambiar el programa informático y éste no podía adaptarse por sí solo a la nueva situación, hubo que buscar una solución de compromiso, que consistió en modificar la trayectoria prevista para minimizar el efecto Doppler. Afortunadamente, este reajuste fue suficiente para evitar que se perdiera la información de la sonda, pero esto indica que los objetivos iniciales de dotar a la cápsula de inteligencia artificial autónoma estaban muy lejos de haberse cumplido. Tampoco la misión *New Horizons*, lanzada en 2006 y que en 2015 ha llegado a Plutón, puede considerarse un ejemplo de inteligencia artificial autónoma, por lo que los grandes proyectos de la NASA de los años ochenta no se han cumplido satisfactoriamente.

3.5 ¿Es posible la inteligencia artificial fuerte? ↑

Como hemos visto, se dice con frecuencia que estamos cerca de conseguir la verdadera *inteligencia artificial*, la de máquinas tan inteligentes (o más) que los seres humanos. *¿Es esto posible, y si lo es, está realmente tan cerca como parecen creer algunos expertos y los medios de comunicación?*

Para responder a esta pregunta hace falta saber algo más que informática, hay que adentrarse en los campos de la biología y de la filosofía. Después de todo, la *inteligencia artificial* es una copia. Existe una *inteligencia natural* que nos sirve de base y punto de comparación. La pregunta anterior puede reformularse así: *¿sabemos lo que es la inteligencia natural, cómo surge y cómo se desarrolla, para poder emularla en nuestras máquinas?* Porque si no sabemos lo que estamos buscando, difícilmente vamos a conseguirlo.

A esta pregunta se le han dado cuatro respuestas filosóficas diferentes e incompatibles (Soler 2013, Polaino 2014):

1. **Dualismo metafísico**: la mente y el cerebro son dos realidades diferentes. La primera es una sustancia espiritual y no espacial, capaz de interaccionar con el cerebro, que es material y espacial. Ambas entidades pueden existir independientemente la una de la otra (Descartes 1647), aunque el cuerpo sin la mente acaba por descomponerse.
2. **Dualismo neurofisiológico**: la mente y el cerebro son diferentes, pero están tan íntimamente unidas que llegan a constituir una unidad (Eccles 1984), son dos estados complementarios y únicos de un mismo organismo (Damasio 1996).
3. **Monismo emergentista**: la mente es un producto evolutivo emergente con auto-organización, que ha surgido como un sistema complejo a partir de sistemas más simples formados por las neuronas (Clayton 1999, 2004, Damasio 1998, Kauffman 2003, Searle 2004). Clayton, por ejemplo, sostiene que las estructuras subyacentes no pueden determinar por completo la evolución de los fenómenos mentales, pero que estos sí pueden influir sobre aquellas.
4. **Monismo reduccionista o funcionalismo biológico**: la mente está totalmente determinada por el cerebro, y este por la red de neuronas que lo constituye. El pensamiento humano es un epifenómeno. La libertad de elección es una ilusión. Somos máquinas programadas (Dennett 1991, Penrose 2004).

Es evidente que los partidarios de la opción número 4 (el monismo reduccionista) creen de forma natural que la inteligencia artificial fuerte debe ser posible, incluso con máquinas como las que ya tenemos, cuando alcancen una potencia y velocidad suficientes. Los partidarios de la opción número 3 (monismo emergentista) tienden a pensar que, para que la inteligencia artificial fuerte sea posible, primero tenemos que cumplir una condición previa: conocer a

fondo el funcionamiento de nuestro cerebro, antes de que seamos capaces de simularlo (Hawkins 2004). Para los que abrazan una de estas dos opciones, la mente no es más que el *software* de nuestro cerebro, el programa que le hace funcionar. Por eso, ambas posturas filosóficas tienden a ser *funcionalistas*.

En cambio, los partidarios de las dos primeras opciones, que piensan que la mente y el cerebro son entes distintos, aunque puedan estar íntimamente relacionados, suelen opinar que crear una inteligencia artificial fuerte no estará nunca a nuestro alcance, porque antes tendríamos que ser capaces de crear mentes, lo que es probablemente imposible, puesto que la mente estaría fuera del alcance de la física, y por tanto de la tecnología.

4 Consideraciones finales ↑

En este artículo hemos revisado las dos acepciones de la Inteligencia Artificial: la débil, que ya está entre nosotros, cuya inteligencia es bastante discutible, pero que ha dado lugar a muchas aplicaciones interesantes; y la fuerte, la verdadera inteligencia, que aún no existe y quizá nunca llegue a existir. Y si existiera, ni siquiera hemos llegado a un acuerdo sobre los medios que podríamos emplear para detectarla.

Siempre es arriesgado predecir el futuro, pero parece claro que muchos de los avances que se anuncian con ligereza como inminentes están aún lejanos. Por ejemplo, no es probable que el objetivo de comunicarse con las máquinas en lenguaje totalmente natural se encuentre siquiera a pocas décadas de distancia. Y, por supuesto, las predicciones de Ray Kurzweil, algunas de las cuales ya han sido refutadas por el paso del tiempo, no van a cumplirse en un futuro inmediato, suponiendo que sean posibles.

No sabemos, ni hay acuerdo entre los expertos, si será factible o no, por medios informáticos, construir inteligencias iguales o superiores a la nuestra, con capacidad de auto-consciencia. Por medios biológicos, es evidente que sí podemos hacerlo, pues una persona puede engendrar hijos con inteligencia igual o superior a la suya. ¿Será posible a través de la tecnología? Y si lo fuese, ¿podemos asegurar que la construcción de estas máquinas no se convertirá en una amenaza para nuestra existencia?

En el estado actual de nuestros conocimientos no tenemos respuesta a esta pregunta. Esta cuestión se sale fuera del campo de la informática y se adentra en el de la filosofía, pues está íntimamente ligado con el problema de la mente y la consciencia humanas, que aún no está resuelto a satisfacción de todos.

5 Bibliografía ↑

Ackerman, Evan. 2014. “Can Winograd Schemas Replace Turing Test for Defining Human-Level AI?” *IEEE Spectrum*, posted 29 Jul 2014, http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/winograd-schemas-replace-turing-test-for-defining-humanlevel-artificial-intelligence.

Alfonseca, Manuel. 1980. “Automatic solution of sorites” *Kybernetes*, Vol.9:1, 37-44. doi: 10.1108/eb005540.

Alfonseca, Manuel. 2015. “Vivir, o el poder de la abducción” http://divulciencia.blogspot.com/2015/10/vivir-o-el-poder-de-la-abduccion.html.

Alfonseca, Manuel; Cebrián, Manuel; Fernández-Anta, Antonio; Coviello, Lorenzo; Abeliuk, Andrés; Rahwan, Iyad. Inédito. “Superintelligence cannot be contained”.

Asimov, Isaac. 1950. *I, robot*. Doubleday, New York.

BBC News Online. 2016. “Artificial Intelligence: Google’s AlphaGo beats Go master Lee Sedol”, 12 March 2016, http://www.bbc.com/news/technology-35785875.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. Springer.

Bostrom, Nick. 2005. “A history of transhumanist thought” *Journal of Evolution and Technology*. http://www.nickbostrom.com/papers/history.pdf.

Bowling, Michael; Burch, Neil; Johanson, Michael; Tammelin, Oskari. 2015. “Heads-up limit hold’em poker is solved” *Science*, 347(6218):145–149.

Campbell, Murray; Hoane, A. Joseph and Feng-hsiung Hsu. 2002. “Deep Blue” *Artificial Intelligence*, 134(1-2):57–83.

Chess News. 2004. http://en.chessbase.com/post/bilbao-man-vs-machine-a-resume.

Chess News. 2005. http://en.chessbase.com/post/8-4-final-score-for-the-machines-what-next-.

Clayton, Philip. 1999. “Neuroscience, the person and God: an emergentist account”. En *Neuroscience & the person. Scientific perspectives on divine action*, ed. Russell, R.J.; Murphy, N.; Meyering, T.; Arbib, M. Vatican Observatory Publications, Vaticano.

Clayton, Philip. 2004. *Mind and Emergence, from Quantum to Consciousness*. Oxford University Press, Oxford.

Clocksin, William F.; Mellish, Christopher S. 1981, 2003. *Programming in Prolog*. Springer, Berlin, New York. ISBN: 978-3-540-00678-7.

Damasio, Antonio. 1996. *El error de Descartes*. Gijalbo-Mondadori, Barcelona.

Damasio, Antonio. 1998. *The feeling of what happens: body and emotion in the making of consciousness*. Harcourt, New York.

Dennett, Daniel. 1991. *Consciousness explained*. The Penguin Press. Existe traducción española: *La conciencia explicada*. Paidos, Barcelona.

Descartes, René. 1647. *La description du corps humain*.

Dickmanns, Ernst Dieter. 2007. *Dynamic Vision for Perception and Control of Motion*. Springer.

Dong Yu, Li Deng. 2015. *Automatic speech recognition*. Springer. ISBN: 978-1-4471-5778-6.

Eccles, J.; Robinson, D.N. 1984. *The wonder of being human. Our brain & our mind*. Free Press, New York.

Faundez Zanuy, M. 2000. *Tratamiento digital de voz e imagen y aplicación a la multimedia*. Marcombo, Barcelona.

Feigenbaum, Edward A.; McCorduck, Pamela. 1983. *The fifth generation*. Addison-Wesley, Reading MA. ISBN: 978-0-201-11519-2.

Ferrucci, David; Brown, Eric; Chu-Carroll, Jennifer; Fan, James; Gondek, David; Kalyanpur, Aditya A.; Lally, Adam; Murdock, J. William; Nyberg, Eric; Prager, John; Schlaefer, Nico; Welty, Chris. 2010. “Building Watson: an overview of the DeepQA project” *AI Magazine*.

Fisher, Adam. 2013. “Inside Google's quest to popularize self-driving cars”. *Popular Science*. September 18 2013. http://www.popsci.com/cars/article/2013-09/google-self-driving-car.

Goldberg, David. 1989. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley, Reading MA. ISBN: 978-0201157673.

Gurney, K. 1997. *An introduction to neural networks*. Routledge, London UK. ISBN: 1-85728-503-4.

Hawkins, Jeff. 2004. *On intelligence*, St.Martin’s Griffin, New York.

Hernandez-Orallo, J; Dowe, D L. 2010. "Measuring Universal Intelligence: Towards an Anytime Intelligence Test". *Artificial Intelligence Journal* 174 (18): 1508–1539, doi:10.1016/j.artint.2010.09.006.

Hutchins, W.J.; Somers, Harold. 1992. *An introduction to machine translation.* Academic Press, London. ISBN: 0-12-362830-X.

IBM Research. 2011. http://researcher.watson.ibm.com/researcher/view_group.php?id=2099. Véase también http://www.research.ibm.com/cognitive-computing.

Jacobs, W.W., *The monkey's paw*, 1902. http://americanliterature.com/author/w-w-jacobs/short-story/the-monkeys-paw.

Jochem, Todd; Pomerleau, Dean; Kumar, Bala; Armstrong, Jeremy. 1995. "PANS: A Portable Navigation Platform". The Robotics Institute. http://www.cs.cmu.edu/afs/cs/usr/tjochem/www/nhaa/navlab5_details.html.

Jurafsky, Daniel; Martin, James H. 2008. *Speech and Language Processing*, Pearson Prentice Hall. ISBN: 978-0-13-187321-6.

Kauffman, S. 2003. *Investigaciones: complejidad, auto-organización y nuevas leyes para una Biología general*. Tusquets. Barcelona.

Kruse, Borgelt, Klawonn, Moewes, Steinbrecher, Held, 2013*. Computational Intelligence: A Methodological Introduction*. Springer. ISBN: 978-144-715012-1.

Kurzweil, Ray. 1990. *The age of intelligent machines*, MIT Press, Cambridge.

Kurzweil, Ray. 1999. *The age of spiritual machines*, Penguin Books, New York.

Kurzweil, Ray. 2005. *The singularity is near*, Viking Books, New York.

Kurzweil, Ray. 2012. *How to create a mind: the secret of human thought revealed*, Viking Books, New York.

Legg, Shane; Hutter, Marcus. 2007. *A Collection of Definitions of Intelligence* (Technical report). IDSIA. arXiv:0706.3639.

López de Mántaras, Ramón. 2013. "Ramón López de Mántaras: Intentar desarrollar una inteligencia artificial igual que la humana es absurdo". CiViCa. http://www.investigadoresyprofesionales.org/drupal/content/ram%C3%B3n-l%C3%B3pez-de-m%C3%A1ntaras-intentar-desarrollar-una-inteligencia-artificial-igual-que-la-humana.

Loveland, Donald W. 1978. *Automated theorem proving: A logical basis*. Fundamental Studies in Computer Science Vol. 6. North-Holland Publishing.

McCarthy, John. 2007. *What is artificial intelligence?* http://www-formal.stanford.edu/jmc/whatisai/whatisai.html.

McCulloch, Warren. Pitts, Walter. 1943. "A Logical Calculus of Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics* 5(4): 115–133. doi: 10.1007/BF02478259.

Manning, Christopher; Schütze, Hinrich. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press. ISBN 978-0-262-13360-9.

Manning, Christopher; Raghavan, Prabhakar; Schütze, Hinrich. 2008. *Introduction to Information Retrieval*. Cambridge University Press. ISBN 978-0-521-86571-5

Minski, Marvin; Papert, Seymour. 1969, 1972. *Perceptrons: An Introduction to Computational Geometry*, The MIT Press, Cambridge MA, ISBN 0-262-63022-2.

Minski, Marvin. 1980. “Decentralized minds”. *Behavioral and brain sciences* 3(3) 439-440. doi: 10.1017/S0140525X00005914.

NASA 2011. *Cassini-Huygens quick facts*. http://saturn.jpl.nasa.gov/mission/quickfacts/. Consultado 2011-08-20.

Nilsson, Nils. 1998. *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann. ISBN 978-1-55860-467-4.

Norris, James R. 1998. *Markov chains*. Cambridge University Press.

Penrose, Roger. 2004. *The Road to Reality*. Random House. Londres.

Polaino, Aquilino. 2014. “¿Ha demostrado la neurociencia que la mente no es más que un subproducto de la materia?”. En *60 preguntas sobre ciencia y fe respondidas por 26 profesores de universidad*, ed. Soler Gil, Francisco José, y Alfonseca, Manuel. Stella Maris, Barcelona.

Popper, Karl. 1962. *La lógica de la investigación científica*, Tecnos.

Powers, David; Turk, Christopher. 1989. *Machine Learning of Natural Language*. Springer-Verlag. ISBN 978-0-387-19557-5.

Rosenblatt, F. 1958. "The Perceptron: A probabilistic model for information storage and organization in the brain". *Psychological Review* 65 (6): 386–408. doi: 10.1037/h0042519.

Russell, Stuart J.; Norvig, Peter. 2003. *Artificial Intelligence: a modern approach*. Prentice Hall, Upper Saddle River. ISBN: 0-13-790395-2.

Schaeffer, J.; Burch, N.; Bjornsson, Y.; Kishimoto, A.; Muller, M.; Lake, R.; Lu, P.; Sutphen, S. 2007. “Checkers Is Solved” *Science*, 317(5844):1518–1522.

Searle, John R. 1980. “Minds, brains and programs”. *Behavioral and mind sciences* 3 (3): 417–457, doi: 10.1017/S0140525X00005756.

Searle, John R. 1999. *Mind, language and society*, Basic Books, New York. ISBN 0-465-04521-9.

Searle, John R. 2004. *Mind. A Brief Introduction*. Oxford University Press. Oxford.

Siddique, Adeli; Nazmul, Hojjat. 2013. *Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing*. John Wiley & Sons.

Soler Gil, Francisco José. 2013. *Mitología materialista de la ciencia*. Ediciones Encuentro, Madrid.

Sutskever, Ilya; Vinyals, Oriol; Le, Quoc V. 2014. *Sequence to Sequence Learning with Neural Networks*. arXiv:1409.3215.

Tesauro, Gerald (1989). “Neurogammon Wins Computer Olympiad”. *Neural Computation* 1 (3): 321–323. doi: 10.1162/neco.1989.1.3.321.

Turing, Alan. 1937. “On Computable Numbers With an Application to the Entscheidungsproblem”. *Proceedings of the London Mathematical Society*, Series 2, Vol.42, 230–265, doi: 10.1112/plms/s2-42.1.230.

Turing, Alan. 1950. “Computing Machinery and Intelligence” *Mind* LIX (236): 433–460. doi:10.1093/mind/LIX.236.433.

United Nations. 2015. *Revision of World Population Prospects*. http://esa.un.org/wpp/Demographic-Profiles/index.shtm.

Weizenbaum, Joseph. 1966. "ELIZA—A computer program for the study of natural language communication between man and machine". *Communications of the ACM* 9 (1): 36–45, doi: 10.1145/365153.365168.

Wiener, Norbert. 1948, 1961. *Cybernetics*, The MIT Press, Cambridge Mass.

Wolfram, Stephen. 2012. “Announcing Wolfram|Alpha Pro”. Wolfram|Alpha Blog, February 8, 2012. http://blog.wolframalpha.com/2012/02/08/announcing-wolframalpha-pro/

6 Cómo Citar ↑

Alfonseca, Manuel. 2016. "Inteligencia artificial". En Diccionario Interdisciplinar Austral, editado por Claudia E. Vanney, Ignacio Silva y Juan F. Franck. URL=http://dia.austral.edu.ar/Inteligencia_artificial

7 Derechos de autor ↑

ISSN: 2524-941X