

¿Cómo medir la **inteligencia** de las **máquinas**?



Carlos Gershenson



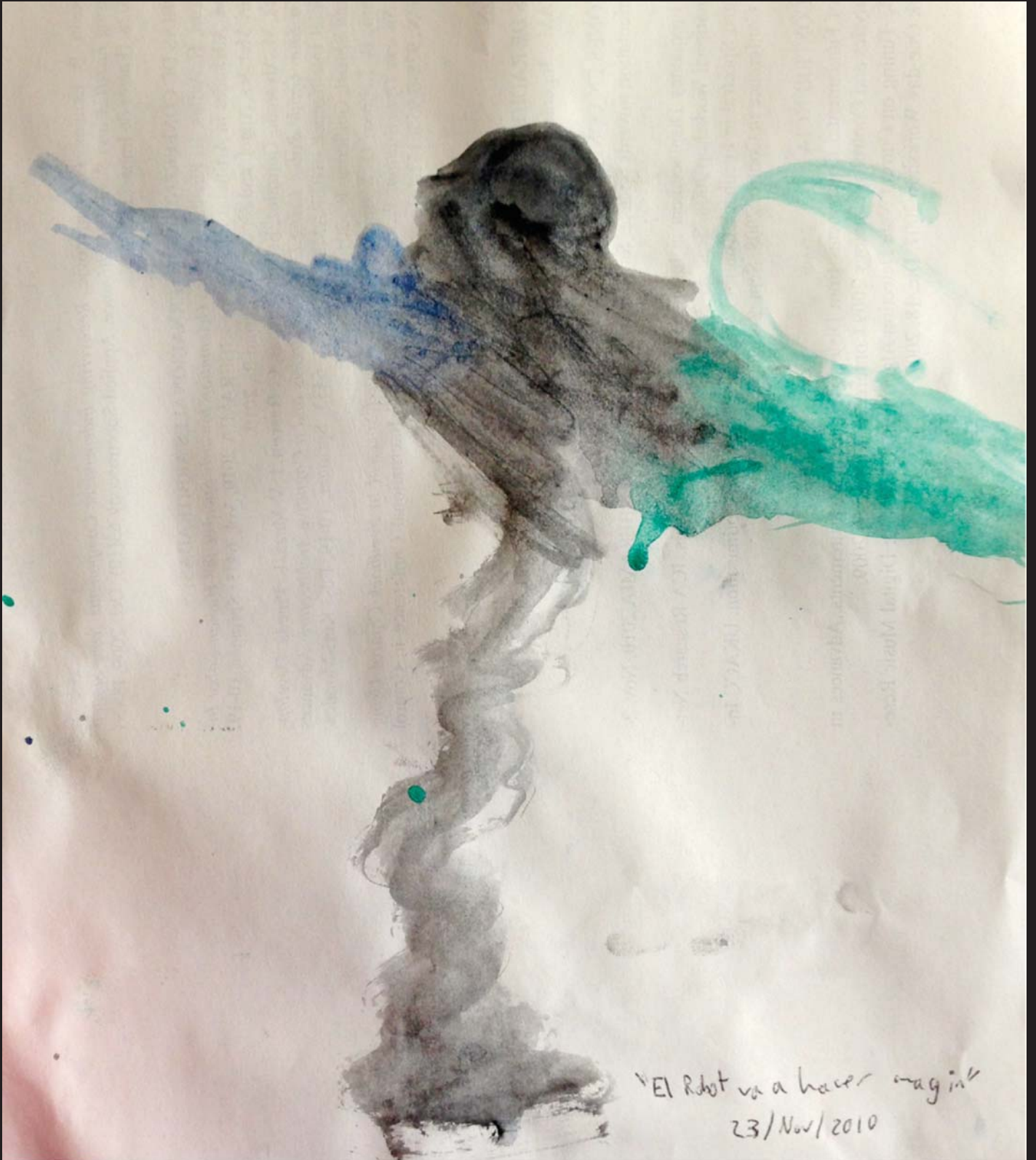
El matemático inglés Alan Turing trató de responder esta pregunta en 1950 proponiendo una prueba que, de ser superada, determinaría si una máquina es inteligente. En la prueba de Turing, una máquina trata de engañar a un examinador para que crea que es una persona. Se podría argumentar que esta prueba no mide tanto la inteligencia en general, sino qué tanto una máquina puede imitar a los humanos. Sin embargo, es una buena base para preguntarnos sobre la naturaleza de la inteligencia y sobre cómo podemos construir sistemas inteligentes.

Introducción

Las computadoras electrónicas, cuyos orígenes datan de la Segunda Guerra Mundial, son uno de los desarrollos tecnológicos más trascendentes de la tecnología. La habilidad de estas máquinas para resolver operaciones aritméticas y ecuaciones mucho más rápidamente que los humanos llevó pronto a la siguiente pregunta: ¿pueden las máquinas ser inteligentes? Es una interrogante difícil desde el principio, ya que todavía no nos ponemos de acuerdo acerca de cómo definir la inteligencia. Y no es tanto que no tengamos una definición: el problema es que tenemos muchas, y podemos juzgar a una máquina (o a una persona) como inteligente con un criterio, pero como tonta con otro. Además, también hay muchas nociones de qué es una máquina, lo cual complica aún más la cuestión.

El primer intento de formalizar la pregunta sobre la inteligencia de las máquinas lo hizo Alan Turing, quien en 1950 publicó un artículo titulado “Computing machinery and intelligence” (Maquinaria computacional e inteligencia) en la revista de psicología y filosofía *Mind* (Mente). En lugar de preguntar si las máquinas podían pensar, propuso una prueba que, de ser pasada por un artefacto, justificaría que se le juzgara como inteligente.

La prueba se basa en el “juego de imitación”, donde un interrogador hace preguntas a un hombre y a una mujer, y trata de adivinar quién es quién. Las preguntas



"El robot va a hacer magia", Ana Gershenson Bazlova. Acuarela sobre papel, 2010.



Alan Turing.



Alan Turing terminando un maratón.

se hacen a distancia, de manera escrita o mediante un intermediario, para que el interrogador no pueda ver ni oír a los jugadores. El objetivo del hombre es engañar al interrogador para que haga la identificación incorrecta y el objetivo de la mujer es ayudar al interrogador. Turing propuso un juego similar, donde un interrogador hace preguntas a un humano y a una máquina. La máquina trata de engañar al interrogador para que crea que es el humano, mientras que el humano trata de ayudar al interrogador para que haga la identificación correcta. La ventaja de esta prueba es que se puede examinar y medir qué tan exitosa es una máquina.

Turing predijo que alrededor del año 2000 una máquina con un gigabyte de memoria podría tener un comportamiento tal que el evaluador promedio juzgaría que la máquina era el humano más de 70% de las veces después de cinco minutos de preguntas.

De hecho, la predicción fue bastante buena. Desde 1991 se ha organizado anualmente el premio Loebner, que galardona a los programas que tengan un mejor desempeño en la prueba de Turing. Ya ha habido varios programas, conocidos como “agentes conversacionales” o “chatbots”, que han logrado engañar a usuarios inexpertos (como lo planteó Turing), pero más usando trucos que exhibiendo comportamiento inteligente. No han logrado engañar a expertos, ya que se sabe cómo hacer preguntas para desenmascarar a un programa. Por ejemplo, el primer ganador del premio Loebner lo consiguió imitando errores de mecanografía, más que imitando inteligencia humana. Es cierto que una máquina se parecerá más a un humano si se equivoca como un humano, pero el hacerlo no nos dice nada sobre la inteligencia, ni en máquinas ni en humanos.

Lo que mide la prueba de Turing es algo más limitado que la muy ambigua y problemática pregunta de si las máquinas pueden pensar, o de si pueden ser inteligentes. Básicamente, trata de comparar las habilidades de una máquina con las de un humano; en este caso particular, para mantener conversaciones.

Sería demasiado arrogante pensar que para que el comportamiento de una máquina sea inteligente éste tenga que ser parecido al del humano, ya que hay comportamientos humanos tontos, al igual que hay comportamientos no humanos que podemos considerar inteligentes.

Como vemos, hay problemas con la prueba de Turing, y en general se encuentran obstáculos similares cuando pretendemos juzgar la inteligencia de animales, plantas, bacterias, extraterrestres (como en el proyecto SETI de búsqueda de inteligencia en otros mundos) y, por supuesto, de humanos.

A pesar de no contar con una definición precisa de inteligencia, podemos decir que desde la publicación del artículo de Turing se han dado avances considerables en la tecnología que han permitido que describir a máquinas como “inteligentes” sea apropiado. Finalmente, si la biología no ha necesitado de una definición precisa de “vida” para alcanzar todos sus logros, no debemos ver la ambigüedad del concepto de “inteligencia” como un obstáculo.

● **Inteligencia artificial**

En 1956 se organizó en Dartmouth, New Hampshire, en Estados Unidos, un congreso donde se acuñó el término “inteligencia artificial” para referirse a máquinas y sistemas capaces de exhibir habilidades cognitivas.

En sus primeros años, la inteligencia artificial se enfocó en problemas que se podían formalizar fácilmente, tales como la prueba de teoremas, juegos de ajedrez y deducciones lógicas. Se han desarrollado “sistemas expertos” en los cuales se usa una “base de conocimientos” para dar respuestas a preguntas usando un “motor de inferencia”. El problema de este enfoque es que es muy limitado para situaciones abiertas, ya que no es posible introducir todas las posibilidades de interacción de antemano, y sus métodos de aprendizaje todavía no han podido adaptarse a ambientes complejos.

En respuesta a estos y otros obstáculos, a fines de la década de los ochenta y principios de los noventa del siglo pasado se propuso una “nueva” inteligencia artificial, enfocándose más en el comportamiento animal que en el razonamiento humano. Este enfoque produjo avances en locomoción de robots, procesamiento de señales y sistemas adaptativos.

En años recientes se siguen practicando –y mezclando– ambos “sabores” de inteligencia artificial: para aplicaciones particulares, importa más la función que el paradigma que se maneje. Y, a final de cuentas, se ha



mostrado que en teoría son equivalentes (Gershenson, 2004, 2010).

Algunos de los múltiples logros de la inteligencia artificial son la automatización de procesos, algoritmos de optimización, algoritmos de búsqueda (por ejemplo, Google), procesamiento de lenguaje natural, visión computacional y generación de leyes a partir de datos (Schmidt y Lipson, 2009).

Todavía hay muchos problemas abiertos en inteligencia artificial, debido a que es un área fértil de investigación. Para compensar los límites de la inteligencia artificial se ha usado la computación humana, también llamada “inteligencia artificial artificial”, donde se usa a humanos para resolver tareas que se les dificultan a las máquinas. Por ejemplo, las pruebas CAPTCHA para reconocer si un usuario no es un programa automatizado (de hecho, una prueba inversa de Turing, ya que nosotros tenemos que convencer a una máquina de que somos humanos), explotan nuestro reconocimiento de patrones para digitalizar libros. En otro ejemplo, Google usa un juego para que usuarios etiqueten imágenes encontradas en la red, lo cual está muy lejos de ser posible para un programa.

● **Muchas pruebas**

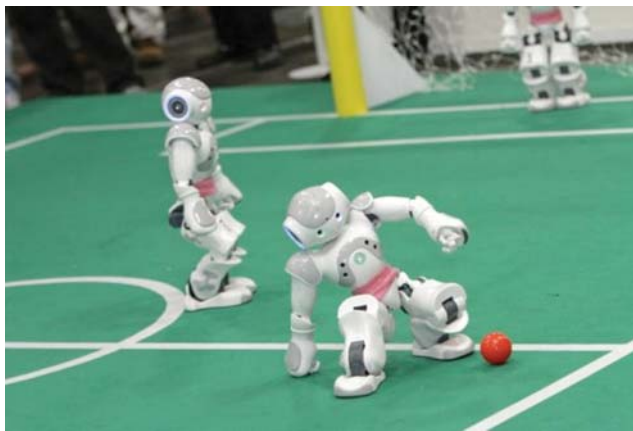
Podemos decir que la prueba de Turing original mide las habilidades de conversación de una máquina. Pero, ¿qué hay de otras habilidades? Podríamos hablar de distintas pruebas para distintas habilidades, no tanto comparando si las máquinas pueden ser tan buenas (o malas) como los humanos, sino compitiendo para ver quién gana.

Algunas computadoras ya juegan mejor que los humanos ajedrez, damas, scrabble, jeopardy y otros juegos. Todavía ganamos en juegos como póker y go, pero es probable que en el futuro cercano este resultado cambie.

Los juegos mencionados anteriormente son mentales, pero también se han propuesto pruebas para otros campos. Por ejemplo, el Reto DARPA para automóviles autónomos (ganado en 2005 por el equipo de la Universidad de Stanford) o una prueba para fútbol. En 1997 se llevó a cabo la primera competencia internacional de robótica RoboCup, con la meta de que en 2050 un equipo de robots humanoides autónomos le gane al equipo campeón de la copa mundial de fútbol. En 2012, RoboCup se realizó en México. Aunque en 15 años ha habido un gran progreso, todavía falta mucho para que los robots bípedos puedan siquiera correr pateando un balón, pero falta también mucho para 2050, así que aún no es posible especular si la meta se cumplirá o no.

Hay muchas áreas en las cuales las máquinas fácilmente superan a los humanos: cálculos matemáticos, manufactura, fuerza, velocidad, etcétera. También hay muchas tareas en las que los humanos superamos fácilmente a las máquinas; el fútbol todavía se halla entre ellas.

En las pruebas consideradas “de inteligencia” en que las máquinas nos ganan, lo hacen de manera muy distinta a nosotros. Se podría argumentar que no nos ayudan a comprender la inteligencia humana, pero también se podría contestar que lo que importa son los



Robots jugando fútbol en la RoboCup (www.robocup.org).

resultados: los aviones no vuelan como aves ni como insectos (que, de hecho, son mucho más eficientes que los aviones), pero si lo que nos interesa es volar, no importa tanto cómo lo logremos. De manera similar, si queremos construir sistemas inteligentes con cierto propósito, no importa si no funcionan de manera similar a los humanos o a otros seres vivos.

● ¿imitación es inteligencia?

Una de las críticas a la inteligencia artificial más discutidas en ciencias cognitivas ha sido el “problema del cuarto chino”, propuesto por John Searle (1980). Plantea imaginar un cuarto con un inglés que no habla chino, pero tiene un manual para manipular símbolos chinos, papel y lápiz. Recibe instrucciones en chino. Consulta el manual, el cual le dice cómo manipular los símbolos recibidos y dar una respuesta en chino también. El inglés da la respuesta correcta, pero no *sabe* qué fue lo que hizo.

Searle argumenta que un programa “inteligente” es similar al inglés en el cuarto chino: puede dar la respuesta correcta (pasar la prueba de Turing), pero no tiene la “intención” de dar esa respuesta, ni entiende qué es lo que hizo. No “piensa realmente”.

Desde un punto de vista pragmático, no importa si una máquina entiende o no, mientras cumpla con el propósito para el cual fue diseñada. Pero desde un punto de vista más filosófico, se ha argumentado que si una máquina no entiende, entonces no podríamos decir que tiene una mente, ni que es inteligente.

Por un lado, podríamos afirmar que nosotros también caemos en el problema del cuarto chino, ya que pensamos sin saber cómo pensamos, además de que la mayoría de nuestras acciones son inconscientes; por otro lado, sí hay muchas explicaciones que podemos dar sobre nuestra mente, hasta cierto nivel. ¿No podría una máquina hacer lo mismo?

Se podría reducir este problema al de cómo se adquiere *significado*. Para que yo pueda tener cierta intención o entender algún concepto, necesito incorporar significados. También una máquina. Pero, ¿dónde están los significados? Siguiendo a Wittgenstein (2009) podemos decir que el *uso* de un símbolo o concepto determina su significado; por lo tanto, si una persona,

un animal o una máquina hace un *uso* adecuado de un símbolo o concepto, entonces afirmaríamos que lo *entienden*. De nuevo regresamos a una perspectiva más pragmática en la que no nos interesa tanto discutir si una máquina o un humano tienen una mente, ya que es algo muy difícil de definir y por tanto de medir, sino que buscamos juzgar si el comportamiento de una persona, animal o máquina satisface las expectativas de un observador para decidir si el comportamiento fue inteligente. Simplemente hay que estar conscientes de que la decisión dependerá no sólo del sistema juzgado, sino de las expectativas del observador que emite el juicio. Por lo tanto, el mismo sistema puede ser juzgado como inteligente o tonto, dependiendo de la prueba y del observador.

¿Qué nos espera?

La prueba de Turing ha sido superada, pero a base de trucos más que de inteligencia. Esto no nos ayudó a entender mejor la inteligencia humana ni a construir sistemas inteligentes, pero desarrollos en las distintas ramas de las ciencias cognitivas y de inteligencia artificial lo están logrando, poco a poco.

Cada vez tenemos sistemas más inteligentes (independientemente de nuestra definición de inteligencia), y cada vez entendemos mejor nuestra mente: cómo es que podemos preguntarnos “¿cómo es que podemos preguntarnos?”.

Al observar el desarrollo de sistemas artificiales inteligentes ha sido común, desde hace décadas, la visión de un futuro donde los humanos nos volvemos obsoletos y las máquinas conquistan el mundo (el escenario de la saga de películas de *Terminator*). Podemos ver que esta idea se debe a una comprensión sesgada del desarrollo de la tecnología. Como mencionamos, aunque hay muchas pruebas en las que las máquinas son mejores que nosotros, hay muchas otras en las que no vemos que las máquinas puedan llegar a estar cerca de nuestras habilidades. Finalmente, no nos sirve de mucho tener una máquina que se comporte exactamente como un humano, teniendo tantos miles de millones en el planeta. Es cierto que cada vez somos más dependientes de nuestra tecnología, pero más que tender hacia una dominación de las máquinas, nos estamos

inclinando hacia una simbiosis, donde podemos ampliar nuestras capacidades gracias a nuestra tecnología, pero, a su vez, nuestra tecnología requiere de nuestras habilidades para poder persistir. Como escribió Turing (1950): “Podemos ver sólo una corta distancia adelante, pero podemos ver ahí que hay bastante por hacer.”

Carlos Gershenson es investigador y jefe del Departamento de Ciencias de la Computación del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas de la Universidad Nacional Autónoma de México (UNAM). Es también investigador asociado en el Centro de Ciencias de la Complejidad de la UNAM. Realizó una estancia posdoctoral en el Instituto de Sistemas Complejos de Nueva Inglaterra; es doctor en Ciencias *summa cum laude* por la Universidad Libre de Bruselas, en Bélgica; tiene una maestría en Sistemas Evolutivos y Adaptativos por la Universidad de Sussex, en Inglaterra; y es ingeniero en Computación por la Fundación Arturo Rosenblueth. Estudió cinco semestres de Filosofía en la UNAM. Tiene una gran variedad de intereses académicos, incluyendo sistemas auto-organizantes, complejidad, evolución, vida artificial, información, cognición, sociedades artificiales y filosofía.
cgg@unam.mx



Lecturas recomendadas

- Gershenson, C. (2004), “Cognitive paradigms: which one is the best?”, *Cognitive systems research*, 5(2):135-156.
- _____(2010), “Computing networks: a general framework to contrast neural and swarm cognitions”, *Paladyn, Journal of behavioral robotics*, 1(2):147-153.
- Schmidt, M. y H. Lipson (2009), “Distilling free-form natural laws from experimental data”, *Science*, 324(5923): 81-85.
- Searle, J. R. (1980), “Minds, brains, and programs”, *Behavioral and brain sciences*, 3(3):417-457.
- Turing, A. M. (1950), “Computing machinery and intelligence”, *Mind*, 59(236):433-460.
- Wittgenstein, L. (2009), *Philosophical investigations*, 4a. edición, Wiley-Blackwell.