

# La noción de inteligencia después de *Computing Machinery and Intelligence*

UNA PERSPECTIVA HISTÓRICA

PEDRO MONTOTO GARCÍA (USC)  
ENRIQUE ALONSO GONZÁLEZ (UAM)

20 DE JUNIO DE 2016



*"Intelligence is what is measured by intelligence tests."*

---

E. Boring, circa 1920, en Legg y Hutter 2007

**Revisar apendice A** **Revisar apendice B** **Revisar capitulo 3**

## Resumen

Este trabajo pretende estudiar en profundidad el concepto de inteligencia que describe el *juego de la imitación*, también conocido como *Test de Turing* en *Computing Machinery and Intelligence* (Turing 1950). Ofrecemos una panorámica histórica de la evolución tecnológica y filosófica que conduce a este experimento y listamos los pros y contras que el mismo tiene para la detección de inteligencia general. Para ello analizamos la propuesta de Turing para artefactos inteligentes y la relacionamos con los avances tecnológicos desde la publicación de dicho artículo. Se presenta como conclusión la caracterización experimental de los Test de Turing y derivados del mismo, así como sus pros y contras, y la necesidad de avanzar hacia un mejor modelo de experimentos. Hemos añadido apéndices relatando aspectos secundarios de la evolución de la maquinaria de cómputo y la psicología que ayudan a comprender el contexto en el que este artículo fue desarrollado y su evolución posterior.

**Palabras clave:** Historia de la Inteligencia Artificial, Filosofía de la Inteligencia Artificial, Alan Turing, Ciencias Cognitivas, Cibernética

## Abstract

**Keywords:** History of Artificial Intelligence, Philosophy of Artificial Intelligence, Alan Turing, Cognitive Sciences, Cybernetics

# Índice general

---

<b>1 Breve historia de la Inteligencia Artificial</b>	<b>1</b>
<b>2 La base física de la IA: ¿Qué entendemos por construir?</b>	<b>7</b>
<b>3 La definición de inteligencia</b>	<b>13</b>
3.1. <i>Computing Machinery and Intelligence</i> . . . . .	14
3.2. Limitaciones del Test de Turing . . . . .	19
<b>4 Conclusiones</b>	<b>25</b>
<b>A Historia del Hardware Computacional</b>	<b>27</b>
<b>B El giro cognitivista</b>	<b>31</b>
<b>Bibliografía</b>	<b>35</b>

# 1 Breve historia de la Inteligencia Artificial

---

*“Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience.”*

---

Common statement with 52 expert signatories, 1997, en Legg y Hutter 2007

Todo desarrollo relativo a la Inteligencia Artificial comienza con una pregunta de apariencia simple, *¿es posible construir algo que pueda pensar?*, o cuestiones anejas como *¿es posible construir algo vivo?* o *¿es posible construir algo que resuelva cualquier problema matemático?*, que en realidad presentan otras muchas cuestiones asociadas. Este tipo de ideas no es en absoluto reciente ni mucho menos. La idea de la Inteligencia Artificial ha existido en diversas formas durante la historia del pensamiento occidental, al menos desde la Grecia clásica, en mitos, leyendas, historias, especulación y autómatas mecánicos y que, a favor o en contra, intentan dar una respuesta final a esta especie de sueño colectivo. Trataremos de ofrecer un recorrido histórico que nos permita entender el contexto en el que Turing 1950 fue concebido. En este primer epígrafe justificamos el interés del tema tratado en el artículo a analizar y esbozamos una historia de la idea de creación de objetos inteligentes.

En las primeras leyendas griegas de las que tenemos constancia, alrededor del siglo V a.C., se tratan entre otros temas las estatuas de Pigmalión traídas a la vida por Afrodita, diosa de la vida y del amor, y la historia de Hefesto, dios de la forja, que era capaz de construir ayudantes dorados para los dioses. De todo esto nos llega noticia a través de la *Política* de Aristóteles, que crea uno de los primeros ejemplos de ciencia ficción político-social, planteando la cuestión de qué ocurriría si tuviésemos *máquinas/autómatas/seres artificiales* inteligentes:

Pues si cada uno de los instrumentos pudiera cumplir por sí mismo su cometido obedeciendo órdenes o anticipándose a ellas, si, como cuentan de las estatuas de Dédalo o de los trípodes de Hefesto, de los que dice el poeta que entraban por sí solos en la asamblea de los dioses las lanzaderas tejieran solas y los plectros tocan la cítara, los constructores no necesitarían ayudantes ni los amos esclavos. (Aristóteles 1988, Aristot. Pol. 1.1253b)

Prosiguiendo, el Talmud, compilado entre los siglos I y VI, habla de *golems* creados de tierra a los que hombres santos y doctos podrían infundir de vida. En el siglo XII el catalán Ramon Llull construye un conjunto de discos de papel que teniendo como base la lógica escolástica y convenientemente combinados y rotados permitirían dirimir cualquier discusión teológica, sistema que llamó *Ars Magna*. En algún momento entre finales del siglo XV ó principios del XVI Leonardo da Vinci crea unos esquemas para un robot-caballero que sería capaz de sentarse, levantarse y mover los brazos manipulado, eso sí, por un humano<sup>1</sup>.

En el siglo XVII se desarrolla, debido al auge del racionalismo y las ideas humanistas, la idea de que todo puede ser explicable mediante métodos mecánicos y matemáticos<sup>2</sup>, incluidos los seres vivos. Se puede decir incluso que hasta ésta época, desde la Grecia clásica, vida e inteligencia sólo podía ser algo otorgado por dioses u otros seres omnipotentes más allá del universo físico, al darse que cualquier creación humana no puede superar una imitación de la vida que la divinidad otorga. Hobbes, en este mismo siglo en su *Leviathan*, contempla la posibilidad de crear un ingenio mecánico que se comporte como un animal, pues todos los órganos para él tienen paralelismos con la mecánica: “el corazón no es más que un muelle, los nervios son cuerdas” y pasajes similares aparecen a lo largo de su obra. Descartes, por el contrario, creía que las máquinas serían incapaces de pensamiento real pues sólo están formadas por materia y sería imposible dotarlas de mente ya que la *res cogitans* es inmaterial. Algunos pensadores creen ver en Leibniz un avance de la Inteligencia Artificial ya que, al igual que Hobbes, concebía que todo lo que hace la mente son computaciones, en *De arte combinatoria*.

En el siglo XVIII Jacques de Vaucanson presenta un autómatas que es capaz de simular un pato vivo en algunos de sus aspectos, construido mediante un armazón metálico adornado con plumas de pato y componentes de relojería y mecánica, que

<sup>1</sup>Éste robot se presentó funcionando en una fiesta de la época en la corte de Venecia en 1495 organizada por Ludovico Sforza, y más recientemente un empresario llamado Mark Elling Rosheim reconstruyó los diseños de Leonardo, probando que los mismos eran sensatos y funcionales.

<sup>2</sup>Verbigracia citamos el título de la publicación de Isaac Newton *Philosophia Naturalis Principia Mathematica* (i.e. Principios matemáticos de la filosofía natural) cuyo tercer tomo lleva por título *El sistema del mundo*

era capaz de comer grano, beber y “digerir” (En realidad el producto de la digestión ya estaba en el interior del pato para la simulación). Jacques de Vaucanson también es el inventor de las primeras tarjetas perforadas entendidas como contenedoras de programas, secuencias de acciones, para entes mecánicos. Estas tarjetas se usarán en la “programación” de telares mecánicos en el siglo XIX y a principios del siglo XX podremos ver ya máquinas calculadoras que permitían hacer operaciones aritméticas usando estas tarjetas como entrada y salida de información. También en el siglo XVIII se creó el “Turco mecánico”, un autómatas que podía jugar al ajedrez como un maestro y que como exhibición del mismo fue enviado de gira por las cortes de Europa de la época retando y ganando a soberanos y estrategas. Más tarde se supo que éste “autómata” debía su genialidad a un maestro de ajedrez humano que se ponía en su interior en cada partida, una maniobra que continúa siendo usada hoy para entrenar o suplir las capacidades que aún no sabemos construir en máquinas.

A partir del siglo XIX comenzamos a ver por doquier obras de teatro, narraciones y películas que hablan del *qué ocurriría* si las máquinas pudiesen pensar o actuar como humanos<sup>3</sup>. Es evidente, por tanto, que muchos de los problemas que plantea el nacimiento de la IA no son nuevos. El siglo XIX ve el renacer de la lógica como disciplina académica en trabajos como *The Laws of Thought* de George Boole o el *Begriffsschrift* de Gottlob Frege, que son el antecesor directo de las ciencias y maquinaria de la computación.

Para la descripción de las condiciones técnicas en las que se gestó (Turing 1950) comenzaremos por el cerebro, órgano que tradicionalmente se considera como fuente de los comportamientos inteligentes.

Luigi Galvani descubrió, en el siglo XVIII, que las señales neuronales son esencialmente de naturaleza eléctrica al experimentar con corrientes eléctricas y músculos de rana. La disciplina que estudia la “maquinaria” del cerebro, la neurociencia, fue inaugurada a principios del siglo XX por Camilo Golgi y Santiago Ramón y Cajal. Camilo Golgi había inventado un método de contraste que permitía distinguir la estructura del tejido cerebral y sus ramificaciones mediante microscopio, mientras que Ramón y Cajal, continuando el uso de éste método creó la teoría de que el tejido cerebral no se componía de una malla de hilos neuronales, sino que dicho tejido estaba formado por células contiguas pero separadas entre sí, las neuronas.

En (McCulloch y Pitts 1943) se crea el primer modelo computacional bioinspirado

---

<sup>3</sup>Como veremos, entre el pensar y el actuar como humanos se puede adoptar una postura completamente funcionalista, como la que adopta Turing, estableciendo una diferencia entre la percepción de un acto inteligente y el proceso hipotético subyacente que lo hace posible, el pensamiento.



de neuronas artificiales y se demuestra cómo podían usarse dichas neuronas artificiales para calcular funciones lógicas básicas: las neuronas reciben un número de señales de otras neuronas y emiten señal si la intensidad total sobrepasa un umbral interno, es decir, las neuronas pueden expresarse mediante una función de enteros en enteros<sup>4</sup>, en concreto del tipo  $f : \{0, 1\}^n \rightarrow \{0, 1\}$ . El propio Turing (Turing 1937) describe un procedimiento para calcular funciones entre enteros, del que hablaremos en la siguiente sección, que McCulloch y Pitts demostraron equivalente al poder de computación de sus neuronas artificiales, avanzando también sistemas de aprendizaje automático.

Dentro del campo psicológico el trabajo de Turing se encuentra como precedente del *funcionalismo* (Levin 2013), en tanto asume que el proceso mental, aunque imposible de observar directamente, es importante para la generación de comportamiento inteligente. El *funcionalismo* es una escuela de pensamiento psicológico que trata de superar el *conductualismo*, para el cual todos los comportamientos conscientes son el resultado de un condicionamiento inconsciente, mediante refuerzo o inhibición de acuerdo al resultado que se ha obtenido después de cierto comportamiento (i.e. comportamientos con resultado positivo para el actor se promueven y viceversa). Antes de la aparición del conductualismo la única forma que se aceptaba para la elaboración de teorías psicológicas era la introspección, sistema que tendía a provocar que los únicos datos que llegaban a publicarse fuesen los que concordaban con las teorías populares en cada grupo de investigación (Russell y Norvig 1994, p.13).

La aparición de grupos de interés para la creación de máquinas inteligentes provocará que se creen grupos de estudio comunes entre psicólogos, neurocientíficos y científicos de la computación, lo que se ha dado en llamar *ciencias cognitivas*. Relatamos su origen en el apéndice B, ya que no es tan relevante para la elaboración del argumento de Turing, pero realmente importante en la evolución posterior como integración de esfuerzos de la IA, la neurociencia y la psicología.

Tenemos por tanto una arquitectura general de la mente, los elementos mínimos del sistema, las neuronas, estudiados en detalle y además una relación clara con la disciplina de la computación y la comunicación<sup>5</sup>, por lo que dentro de la interpretación mecanicista habitual en la época sería suficiente estudiar sus relaciones y establecer un

---

<sup>4</sup>Una crítica a este sistema (Penrose 2006) afirma que las señales de entrada y salida del sistema de una neurona pueden ser binarias, pero es necesario un sistema en el que se apliquen efectos cuánticos para poder simular completamente ésta función entre conjuntos binarios.

<sup>5</sup>En concreto, el hecho de que ambos sistemas, neuronas y chips electrónicos, estén basados en señales eléctricas y binarias. La publicación en 1949 de *A mathematical theory of Communication* por Warren Weaver, fundador de la cibernética, y Claude Shannon, que inaugura la cuantificación (i.e. la reducción a señales cuantizadas o discretas, que en última instancia pueden ser binarias) de la información contenida en un mensaje, es clave para entender esta relación.

modelo matemático para obtener un sistema que nos permitiese simular el objeto de estudio. Es evidente, en retrospectiva, que ni el conocimiento ni la tecnología de la época eran suficientes para la simulación de un cerebro o de un sistema cognitivo completo, sin embargo la confluencia de estas ideas ha dado fruto en numerosas tecnologías y nuevos campos de estudio.

En definitiva la IA trata de *construir y estudiar* los mecanismos que producen, por un lado, comportamientos o pensamiento de tal forma que se alineen, por otro lado, con unos criterios bien de racionalidad ideal o bien de semejanza con lo humano (Russell y Norvig 1994, p.5), tal que así:

Objetivo	Pensar	Actuar
Racional	Leyes del Pensamiento	Agentes Racionales
Humano	Modelado Cognitivo	Test de Turing

Así, un sistema capaz de *Pensar Racionalmente* debería ser capaz de realizar inferencias correctas a partir de datos dados de acuerdo a sistemas de lógica proposicional, i.e. inferencias inescapables de acuerdo a las supuestas leyes del pensamiento, en el estilo inaugurado por Boole. Un sistema capaz de *Actuar Racionalmente* elabora en éstas leyes del pensamiento un procedimiento de decisión que le permite en toda situación actuar de tal manera que la situación planteada y la decisión tomada deben estar en una relación de consecuencia lógica desde unos principios de decisión racional. Sin embargo, las decisiones tomadas por humanos suelen tener desviaciones e inconsistencias respecto a lo racional (Tversky y Kahneman 1981), por lo tanto presuponemos que los *Pensamientos Humanos* no siguen completamente las leyes de la lógica y es necesario establecer un modelo que los describa completamente. Como es imposible, por el momento, observar directamente los sistemas que dan lugar a las decisiones, se hace imperativo el establecimiento de técnicas que permitan averiguar la relación entre los actos observados y el pensamiento que suponemos hay detrás de los mismos. El Test de Turing se encuadra en la observación de dichos actos y, aunque supone que existe una relación entre pensamiento y acto, no nos permite investigar en profundidad la estructura del pensamiento, como veremos en el apartado 3.

Siendo los humanos los únicos seres con los cuales actualmente podemos compartir pensamientos a través del lenguaje no llega a ser una sorpresa que éste sea el único camino que podemos tomar para la detección de Inteligencia Artificial: un constructo (concepto que exploraremos en el apartado 2) es inteligente si los humanos lo designamos como inteligente en la realización de una tarea de forma relativa a la realización de la misma por parte de un humano, siendo las descripciones de la racionalidad una

idealización del proceso de toma de decisiones humano. O, dicho de otra forma, no podemos afirmar ni negar que exista inteligencia detrás de las acciones humanas, pero sí podemos afirmar que existen comportamientos designados como inteligentes. Por tanto, creemos que la caracterización de inteligencia como una cualidad ideal es contraproducente para los objetivos de la Inteligencia Artificial, siendo únicamente identificables los actos inteligentes para un modelo mental subjetivo de la realidad que un observador de dicho acto posee. Se puede decir, resumiendo, que a falta de una descripción de la inteligencia en sí ésta es esencialmente *performativa*. Ahondaremos en este tema en el apartado 3, ya que es fundamental para entender el Test de Turing.

## 2 La base física de la IA:

### ¿Qué entendemos por construir?

---

*“Intelligence is the ability to use optimally limited resources – including time – to achieve goals.”*

---

Ray Kurzweil, 2000, en (Legg y Hutter 2007)

Para la construcción de inteligencia precisamos, por tanto, definir qué entendemos por construcción. Esto presenta un número de problemas, como veremos, que no entran por completo en el alcance de este trabajo, pero creemos que conviene reflejar el cambio de concepto que se ha dado en el último siglo, aunque los grupos de tecnólogos y científicos (ingenieros, matemáticos, biólogos, psicólogos y científicos sociales) que trabajan en Inteligencia Artificial no suelen tenerlo en cuenta.

Normalmente definimos constructo o artefacto (Franssen, Lokhorst y Poel 2013; Hilpinen 2011) como una entidad en la que algunas o todas sus propiedades preexisten en la intencionalidad de un autor. Restringiremos la definición de entidad a que se trata de un objeto material limitado en el espacio que podemos designar con un nombre, aunque que es un problema mucho más amplio que el alcance de este trabajo, perteneciente a las disciplinas de la metafísica y la epistemología. Por tanto, una entidad dada es un artefacto si éstas propiedades existen en la descripción dada desde la intención del autor y son aceptadas como válidas en la descripción de lo construido por el autor, i.e. el autor determina unas propiedades necesarias desde su intención al crear el objeto, que guiarán el proceso de construcción, y valida que dicho objeto posee las características que se esperaba del mismo.

De acuerdo con la terminología relativa a la Filosofía de la Tecnología, “por *tecnología* se entiende un conjunto de conocimientos de *base científica* que permiten describir,

explicar, diseñar y aplicar soluciones técnicas a problemas prácticos de forma sistemática y racional”(Quintanilla 2000, p. 2). Al existir el objetivo de resolver un problema práctico se entiende que la intencionalidad de dicha tecnología es resolver el citado problema, por lo tanto se define la Inteligencia Artificial como una ciencia cognitiva (ver apéndice B) cuyo objeto de estudio son los sistemas que generan comportamientos inteligentes (por ejemplo, los seres humanos). La Inteligencia Artificial obtiene así conocimiento científico que se emplea en crear una *tecnología* asociada con el objetivo de crear sistemas técnicos o artefactos que puedan comportarse de forma inteligente, tecnología que en general se conoce también como Inteligencia Artificial.

Ya desde los tiempos de Aristóteles, en su *Física* se intuye esta definición en la diferencia entre los “productos naturales” que se generan por sus propios impulsos internos mientras que los “productos artificiales” precisan de una intencionalidad humana. Avicena criticaba en la edad media que la alquimia jamás podría conseguir “sustancias genuínas” como las presentes en la naturaleza precisamente por ser un constructo con intencionalidad humana. En este caso, para mantener el debate cerrado en torno a lo que nos proponemos definir, i.e. lo que es un constructo idóneo para la Inteligencia Artificial, asumiremos que no existe autor en los entes naturales y que son autogenerados en cierta medida por sí mismos o sus predecesores en el tiempo.

En consecuencia, la diferenciación natural-artificial es problemática. Los avances más recientes en biología, química e Inteligencia Artificial y las promesas de avances futuros nos llevan en última instancia que lo “artificial” podría diferir sólo de lo “natural” en una cuestión de proceso, lo que podría convertir la dicotomía “natural”-“artificial” en vacua: Si no podemos distinguir natural de artificial sin conocer el proceso que ha generado la entidad en cuestión y las entidades generadas por procesos “naturales” o “artificiales” son indistinguibles, observar dichas entidades antes de conocer el proceso impide distinguir cuál es “natural” o “artificial” a menos que los etiquetemos arbitrariamente.

Distinciones aún más sutiles agudizan el problema: ¿una persona nacida por fecundación in-vitro o cesárea ha nacido por un proceso natural o artificial?, ¿podemos considerar las personas que demuestren inteligencia y hayan nacido mediante métodos que aparentemente son *no naturales* Inteligencia Artificial?

La razón de esta problemática es que las cualidades del objeto que lo hacen “natural” o “artificial” no son inherentes al objeto una vez la técnica de imitación ha avanzado lo suficiente. Este hecho será de especial relevancia cuando comentemos el juego de la imitación de Turing, siendo además una constante en el progreso de la IA el hecho de que tareas que son realizables por computadores y no por humanos pierden en ese

momento, al ser percibidos por parte de la sociedad, su condición de actos inteligentes<sup>1</sup>.

Por tanto, la distinción “inteligencia natural”-“inteligencia artificial” sólo podría darse en dos casos, al tratarse de una cualidad sólo observable en su manifestación como actos inteligentes: o bien se conoce el proceso de creación de ambas entidades antes de su evaluación como inteligentes o bien se conoce que ambos tienen un asentamiento material diferente, uno de los cuales ha sido identificado como natural previamente (i.e. vida basada en el carbono como natural versus chips de silicio como artificial).

Conviene resaltar que todas las referencias que se han podido compilar para este trabajo no adoptan una postura respecto a las bases filosóficas de la distinción natural-artificial o de la teleología de los componentes de la inteligencia. La mayoría de la literatura se limita a presentar diferentes modelos matemáticos que suelen tomar como base la máquina de Turing, las redes neuronales artificiales, aunque otros llegan a argüir que es necesaria una imitación matemática de los procesos biológicos involucrados en las emociones, normalmente se trata de una lista de tipos de procesos matemáticos, es decir, herramientas para la imitación de la inteligencia según la definición de ésta que se dé.

La convención moderna del constructo computacional, que se instancia físicamente en un soporte tecnológico, y sirve para el soporte físico de la Inteligencia Artificial es la Turing-completitud. No nos detendremos aquí a dar una descripción completa de lo que es la máquina de Turing (o el equivalente cálculo lambda de Church, u otros), sirva decir que es el constructo conceptual en el que se basa la operación de la maquinaria de computación general<sup>2</sup>. En su contexto histórico, 1928, tres años antes de la primera publicación de Gödel sobre el segundo problema de Hilbert y, como continuación del programa formalista, Hilbert elabora una nueva lista de problemas para proponer a la comunidad matemática. *Entscheidungsproblem*, como se conoce por su nombre alemán, que en castellano sería “problema de la decisión”, consiste en dar un procedimiento finito para cualquier fórmula de la lógica de primer orden que nos diga si ésta es un teorema o no. Esta idea es heredera directa del décimo problema de Hilbert postulado en 1900, que consistía en similarmente dar un método de pasos finitos para resolver cualquier ecuación diofántica y en una perspectiva histórica también descende de la filosofía leibniziana, el *Calculus*!, y de una cierta estética constante derivada de las disciplinas deductivas<sup>3</sup>.

<sup>1</sup>Como se puede ver en el *What Computers Still Can't Do* de Hubert Dreyfus

<sup>2</sup>Sobre la teoría matemática que soporta esta máquina conceptual puede leerse en (Hopcroft, Motwani y Ullman 2006; Turing 1937).

<sup>3</sup>Las afirmaciones sobre la *elegancia* matemática son ciertamente una constante en la historia de la humanidad, desde Pitágoras a nuestros días, por lo que podemos considerar que todas éstos

En 1936, cinco años después de *On formally undecidable propositions* se da un nuevo golpe a este formato de modernismo llamado formalismo o positivismo. Alan Turing publica *On computable numbers, with an application to the Entscheidungsproblem*; en el que formaliza una idea intuitiva de máquina capaz de realizar cálculos <sup>4</sup>, demostrando que el Entscheidungsproblem es irresoluble mediante estas máquinas; y Alonzo Church publica *A note on the Entscheidungsproblem*; en el que demuestra que el Entscheidungsproblem es también irresoluble mediante las funciones lambda-definibles (o cálculo lambda) que había estado desarrollando con su estudiante Stephen Kleene. Se considera hoy, dado que Kleene demostró que estas clases de funciones son equivalentes; i.e. toda función lambda-definible puede transformarse mediante un procedimiento finito en una función Turing computable equivalente y viceversa; y que no se ha podido describir ningún algoritmo que no sea describible en estos términos, que esta es la noción formal de computabilidad mediante un procedimiento finito más aproximada a la idea intuitiva de computabilidad: la tesis Church-Turing.

La máquina de Turing intuitiva, desprovista del aparataje formal matemático, consiste en una cinta en la que se pueden escribir y sobrescribir símbolos, teniendo capacidad infinita numerable para símbolos y siendo el símbolo la unidad de lectura y escritura, y un dispositivo de máquina de estados que define una función parcial de aplicación continua hasta que se alcanza un estado de parada. Esta función nos dice que si la máquina se encuentra en un estado  $x_t$  y lee en la cinta un símbolo  $s_i$ : a) escribirá en esa misma posición un símbolo  $s_o$  (que puede ser igual a  $s_i$ ), b) se moverá una posición adelante o atrás, y c) se moverá al estado  $x_{t+1}$ , o se quedará en  $x_t$  según el “programa” descrito por la función. La definición de función recursiva toma pues, la forma de esta máquina de estados con memoria sobrescribible.

Pasamos ahora a dar una descripción de los soportes tecnológicos que se han tenido en cuenta para la Inteligencia Artificial, lo que en la jerga del ingeniero en computación sería la infraestructura, y que se compone normalmente de cuatro elementos que establecen abstracciones en el elemento anterior dando lugar a la posibilidad de construir elementos más complejos:

---

descubrimientos están ligados por las mismas aspiraciones estéticas. Citando a Russell: “Mathematics, rightly viewed, possesses not only truth, but supreme beauty—a beauty cold and austere, like that of sculpture, without appeal to any part of our weaker nature, without the gorgeous trappings of painting or music, yet sublimely pure, and capable of a stern perfection such as only the greatest art can show.” (Russell 1919, p.61)

<sup>4</sup>Que Turing usará en sus implementaciones de las *Bombes*, unas de las primeras máquinas de cálculo eléctricas que fueron usadas para acelerar el descifrado de mensajes alemanes en la Segunda Guerra Mundial

- En el núcleo del artefacto tenemos el soporte físico de cómputo: hablamos generalmente de una máquina de cómputo de propósito general con procesador aritmético-lógico y memoria, que contendrá datos y programa de acuerdo a la arquitectura von Neumann, o más recientemente, de varias de éstas máquinas conectadas mediante una red de datos (llamado procesamiento distribuido) o funcionando sobre una memoria compartida (llamado procesamiento paralelo).
- Éste elemento se suele cargar con un sistema operativo que establece una capa de abstracción sobre los elementos físicos para que los programadores puedan usar todos los componentes de forma simple y eficiente, ya que se encarga también de otorgar el control de recursos físicos a programas.
- Sobre los sistemas operativos se opera mediante lenguajes de programación, que consisten en abstracciones sobre los sistemas del mismo que alcanzan al sistema físico de cómputo y que al mismo tiempo ofrecen una manera de expresar algoritmos. Cada lenguaje ofrece diferentes estructuras computacionales que los hacen más adecuados para expresar un tipo de algoritmo u otro, aunque todos suelen tener la misma capacidad teórica, esto es, la Turing-completitud.
- En última instancia tenemos los algoritmos y estructuras de datos que nos permiten resolver el problema que estemos tratando. Ésta es la parte esencial del sistema que identificamos como Inteligencia Artificial, ya que todos los sistemas actuales poseen pequeñas variaciones de los sistemas precedentes siendo los cambios en algoritmos los más relevantes hoy en día.

Cada una de estas partes puede considerarse como un problema de ingeniería en sí mismo, lo que ayuda a la simplificación de su resolución en conjunto, y a la división en tareas de los grupos de investigación, puesto que en las últimas décadas se ha pasado de la investigación general en computación a la especialización cada vez más granular en cada uno de estos campos. Por supuesto para llegar a estos elementos desde la concepción simple de las bombas, que eran computadores contruídos con un propósito definido y limitado (i.e. no programables y no Turing-completos), hubo una evolución de las máquinas de cómputo, que relatamos en el apéndiceA.

Hoy en día se están empezando a estudiar tecnologías prometedoras como la expansión de la tecnología electrónica mediante elementos ópticos, bajo la promesa de más velocidad de cómputo al ser las interacciones entre fotones más rápidas que las interacciones entre electrones, en la computación cuántica, que utiliza los fenómenos a nivel de partícula atómica para implementar puertas lógicas en una variante de la lógica



multivalorada denominada lógica cuántica que permitiría la paralelización masiva de las operaciones atómicas, la computación basada en ADN y biología molecular, que basa su implementación de puertas lógicas en interacciones químicas de proteínas y ácidos o intercambios químicos entre células, como en las sinapsis neuronales. Todo ello nos conduce a pensar que estas tecnologías tienen una relación inherente con la base física o biológica de los seres vivos, y aunque no resulten productivas para los problemas de computación actuales <sup>5</sup> podrían darnos esa capacidad de imitación a todos los niveles que mencionamos al principio del capítulo.

---

<sup>5</sup>Por ejemplo, la computación basada en elementos biológicos resulta más lenta que su equivalente electrónico pero permite el procesamiento paralelo en una escala mucho mayor, lo que puede ser también una propiedad de la computación cuántica según la implementación que siga

### 3 La definición de inteligencia

---

“[*Intelligence is*] Achieving complex goals in complex environments”

---

B. Goertzel, 2006, en Legg y Hutter 2007

La definición de lo que es inteligencia es naturalmente un apartado importante, aunque de difícil solución en el campo de la Filosofía de la Inteligencia Artificial. Se puede ver un pequeño compendio de definiciones históricas de inteligencia en diferentes ramas de las ciencias cognitivas en (Legg y Hutter 2007). En primer lugar, trataremos de diferenciar la “inteligencia” en sí que nos parece de un platonismo peligroso al declarar que ésta existe como entidad al no ser designable lingüísticamente en el universo físico, y reemplazarla por “comportamientos inteligentes” o “tareas para las que es necesaria inteligencia”. “Inteligencia”, entendemos pues, no es predicable directamente de agentes, sino de actuaciones que califican a dicho agente como inteligente.

Se afirma normalmente que ser inteligente es un atributo humano, aunque si tomamos ciertas definiciones de inteligencia como válidas podemos atribuir inteligencia a todos los seres vivos o incluso a seres carentes de vida y capacidad de acción como piedras<sup>1</sup>, que, intuitivamente no parecen tener ninguno de los atributos necesarios para la inteligencia. Es por tanto un tema que merece ser tratado con cuidado.

Entendemos por tanto que hemos de dar una definición de mente o consciencia desde la que argumentaremos en este trabajo. Decimos que la mente al igual que la consciencia es una propiedad emergente de los sistemas físicos que se manifiesta en los comportamientos inteligentes, pero que carece de sentido separada de éstos. Es decir, los comportamientos inteligentes son la manifestación de lo que podemos llamar

---

<sup>1</sup>En el caso que nos ocupa podemos decir que si aceptamos “*Inteligencia es* la capacidad de mantener una existencia continuada en el tiempo”, ésta definición no especifica que para tener inteligencia ha de ser un ser vivo y si lo especificase ciertamente excluiría la maquinaria de computación. Test como el de Gunderson, que veremos más adelante, entran en este caso también.

inteligencia, mente o consciencia, pero no existen comportamientos inteligentes no físicos, de la misma forma que un algoritmo anotado en un papel o cargado en una tarjeta de memoria electrónica no puede generar comportamientos inteligentes a no ser que haya un actuador que lo interprete y lo haga actuar sobre una parte del universo físico. Los algoritmos o la mente *per se* no poseen la capacidad de generar ni explicar comportamientos inteligentes a menos que estén activos sobre una base material, por tanto, no discutiremos a favor o en contra de la existencia material o inmaterial de la mente: son abstracciones que carecen de valor en este trabajo.

En última instancia, consideramos que mente o consciencia es una abreviación para referirnos a éstos comportamientos inteligentes, que en realidad no aporta valor adicional al debate sobre la IA o su entendimiento. Lo que llamamos intencionalidad de los comportamientos proviene de la relación del sistema con el entorno, es decir, la inteligencia sin un entorno sobre el que actuar de forma inteligente no existe<sup>2</sup>. Aceptamos la crítica de Lakoff a la cognición incorpórea y aceptamos que ésta es imposible, pero no creemos que haya suficientes diferencias relevantes entre un programa dotado de mecanismos de evaluación de su entorno y una persona cuyos mecanismos de percepción están en un cuerpo, afectando a sus procesos de cognición y viceversa (Lakoff 1987; Lakoff 2012). Incluso en el caso de que dichos sistemas corpóreos sean virtuales, por ejemplo en el experimento mental de los cerebros en tarros, y simulados existe la necesidad de un soporte físico, sea éste silicio o carbono, para los mismos.

Entendemos que para que un comportamiento pueda considerarse inteligente debe existir posibilidad de error, entendido como una desviación entre la intencionalidad anterior al acto y el estado del mundo después del mismo, y el agente que realiza dicho comportamiento debe ser capaz de detectar esta divergencia en algún momento y corregirla. Para ello el agente debe ser capaz de continuarse autónomamente durante el tiempo suficiente o crear sucesores de sí mismo. En el segundo caso la inteligencia se daría por selección natural, dado que el entorno condicionaría la supervivencia únicamente de los individuos adecuadamente inteligentes.

### 3.1. *Computing Machinery and Intelligence*

El *paper* fundamental que da título a este trabajo e inaugura formalmente la investigación hacia comportamientos inteligentes y vida artificial por parte de los computólogos e ingenieros en computación es *Computing Machinery and Intelligence* de Alan Turing,

---

<sup>2</sup>Es más, tal y como muestran (Tversky y Kahneman 1981) el modo en que las opciones de decisión son presentadas influye de forma determinante en la elección realizada.

publicado en 1950. Es éste el primer momento en que puede verse una definición, aunque no formal, sí observable y reproducible de lo que es un comportamiento inteligente. A medida que la Inteligencia Artificial ha ido enfocándose en aplicaciones concretas como la creación de algoritmos para diferentes análisis de datos, el *paper* de Turing ha perdido relevancia en éste campo, mientras que en disciplinas en las que el sujeto de la discusión es propiamente la mente humana como la psicología y la filosofía, éste sigue estando en el punto de mira. Si bien el paper ha perdido relevancia en las ciencias computacionales debido a su enfoque en la resolución de problemas concretos no deja de ser sorprendente lo precisas que son las intuiciones de arquitectura y aumento en capacidad computacional que predice Turing.

En resumidas cuentas, Turing describe la máquina formal de una manera breve e intuitiva<sup>3</sup> y enuncia las reglas del juego que permitiría descubrir si una máquina de éste tipo cargada con cierto programa está pensando o no. Éste juego debe prepararse de tal manera de que una mujer, un hombre y un interrogador que puede hacer preguntas a ambos. Es conveniente que la sala de testeo separe a los sujetos par que no puedan verse entre ellos pero haya canales de comunicación claros, preferentemente mediante texto, para que sea imposible inferir la cualidad humana o no del estilo comunicativo o la voz de cada sujeto.

Turing plantea la ejecución de tal forma que el interrogador debe adivinar cuál de los dos sujetos es el hombre y cual la mujer, siendo la tarea del hombre convencer al interrogador de que es la mujer y la de la mujer ayudarle a hacer la identificación correcta<sup>4</sup>. Seguidamente pasa a considerar qué ocurriría si sustituyésemos al hombre por una máquina capaz de realizar conversaciones imitando a humanos. El valor del test en sí, por tanto, no es la definición de inteligencia que usa, sino la identificación de inteligencia con capacidad de realizar actos inteligentes a nivel humano.

Hay un punto que suele escapar a las discusiones sobre este paper sobre el que nos gustaría llamar la atención y es que (Turing 1950) explícitamente dice

“Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, *Can machines think?*”

---

<sup>3</sup>Una definición que hemos ampliado y modernizado en el capítulo 2

<sup>4</sup>Este hecho ha sido analizado por diversos autores, como David Leavitt en *El hombre que sabía demasiado*, como muestra de sexismo o desconfianza para con las mujeres por parte de Turing. Si bien es un tema interesante, no consideramos que sea relevante en el legado del paper de Turing, por lo que no nos detendremos en consideraciones al respecto.

Y es que el juego no trata de identificar siempre a la máquina como humana sino que los errores de identificación de un humano como mujer u hombre sean estadísticamente parecidos a los que se cometan en la identificación como hombre o máquina. Es decir, Turing acepta el error en la identificación y lo hace parte de la prueba ya que es consciente de que una característica fundamental del pensamiento humano es que éste no es exacto y, además, es propenso a errores en múltiples niveles<sup>5</sup>.

Turing pasa a describir las máquinas que podrían ser programadas para participar en el test, que se corresponden con el nivel máquina de la arquitectura de computador digital que describimos en el capítulo anterior, es decir, los niveles de abstracción, como ya hemos explicado, facilitan la implementación del algoritmo proveyendo abstracciones sobre el “bare metal”, la máquina básica de cómputo, pero no pueden aumentar las capacidades teóricas de la máquina. De la misma manera que al realizar deducciones matemáticas no partimos simplemente del concepto de número y nuestra demostración debe partir de ahí siempre, se toman generalizaciones dadas, por ejemplo, al inferir hechos sobre triángulos se recurre en primera instancia a la trigonometría y no se deduce, de nuevo, la trigonometría de los axiomas de la geometría. Turing indica claramente que lo relevante en éste constructo no es la máquina en sí, sino el algoritmo que produce los comportamientos inteligentes.

Turing comenta una serie de 9 objeciones que impedirían la construcción de dicha máquina, que aquí comentamos desde una perspectiva generalizadora, actualizando las críticas a dichas objeciones:

**1, 2, 4, 5, 6, 9: La máquina no puede ser humana** Agrupamos 1) Las máquinas carecen de alma, 2) Las consecuencias de la máquina inteligente son terribles ya que el hombre no debe ser sobrepasado, 4) Las máquinas no pueden tener consciencia ni sensibilidad, 5) Las máquinas no pueden X, siendo X una cualidad humana como “amar, aprender, ser el sujeto de su propio pensamiento, tener errores, etc.”, 6) Ada Lovelace, en sus notas sobre el motor diferencial de Babbage, dice que la máquina “no puede hacer otra cosa que aquella que sepamos ordenarle cómo hacer, mas no tiene pretensión de generar nada por sí misma”, es decir, carecen de creatividad y 9) Las máquinas no pueden tener Percepción Extra-Sensorial, como equivalentes al basarse todas ellas en la idea de que algo construido “no puede tener características humanas”. Consideramos 1, 2, 4 y 9 no afectan al test al ser

---

<sup>5</sup>La pregunta de si un humano no pasa el test deja o no de ser inteligente queda implícita, por ello se hace ésta distinción estadística.

éste funcionalista<sup>6</sup> y las críticas estar fundamentadas en supuestas propiedades de la esencia humana cuya existencia no ha sido probada, o de la estructura interna de la mente. 5 y 6 son dudables ya que también carecen de justificación, basándose en los prejuicios del criticante, siendo “al estar programadas para el test, deberían ser capaces de simular X” una contra-crítica equivalente, que es la que de hecho Turing parece esgrimir.

**3, 7: Objeciones matemáticas** Al ser la base de la máquina inteligente propuesta por Turing un constructo capaz de realizar matemática discreta, surgen las siguientes objeciones: 3) La máquina no puede tratar problemas indecidibles de la lógica al estar basada en ella (i.e. preguntada por el problema de la parada la máquina daría una respuesta errónea o entraría en un bucle infinito), 7) La máquina es un constructo discreto mientras que el cerebro humano es continuo. Turing considera que no existe demostración de que las limitaciones de 3 no ocurran también en la mente humana y, que si una máquina da una respuesta errónea a estos problemas no es relevante puesto que la mayoría de humanos también la daría. Como respuestas a 7 se dan que los ajustes discretos a funciones continuas podrían dar resultados suficientemente similares para las preguntas formuladas en el juego, y además, como ya hemos visto las neuronas operan fundamentalmente en términos discretos.

**8: El comportamiento humano es arbitrario** Ésta objeción consiste en que no existe un procedimiento finito, algoritmo, para calcular o predecir el comportamiento de un sujeto humano. Se puede contraargumentar que no existe *aún*, como no existe una descripción completa de la física etc. o bien se puede argumentar que dado un programa que responde a un número de dieciséis cifras con otro arbitrariamente uno podría similarmente concluir que el programa es humano puesto que exhibe arbitrariedad y no hay reglas claras de su funcionamiento, aunque sabemos que éstas están en su código. Aún así consideramos que es una crítica relevante al funcionamiento del test, puesto que un humano puede identificar a la máquina como humana mientras que otro puede no hacerlo de forma arbitraria.

En la parte final del paper se describe el proceso de aprendizaje máquina, que ha dado un campo propiamente establecido en la inteligencia artificial, y que consiste en hipotetizar que la mente del niño es sustancialmente más simple que la del adulto,

---

<sup>6</sup>Una de las pocas alternativas a 4, la máquina no puede ser consciente, es caer en el solipsismo, de hecho.

siendo la mente del adulto la mente del niño más experiencias de aprendizaje y, esto siendo la clave del constructo, la capacidad de generar ideas secundarias y terciarias a partir de experiencias o ideas inyectadas. Lo que Turing denomina, por analogía con los materiales atómicos, una *mente supercrítica*, en las que la inyección de material nuevo, i.e. ideas y experiencias, provoca una reacción que conduce a ideas nuevas: es decir, capacidad creativa y arbitrariedad, de la misma forma que los humanos generarían conceptos nuevos para ellos mismos o incluso para otros humanos. Por la misma analogía, Turing describe las mentes animales como *mentes subcríticas* e incapaces de creatividad o generación de ideas a partir de ideas. Turing presupone que un método aleatorizado puede ser mejor para la generación de este tipo de aprendizaje, ya que no cree ver reglas en la generación de éste tipo de conocimiento: todo ésto es también la base de otro campo del conocimiento, el razonamiento inductivo.

Existen reformulaciones posteriores del TT que admiten errores y la posibilidad de que la identificación de comportamiento inteligente se haga por un juicio grupal, como el usado en el premio Loebner<sup>7</sup>, en el que la identificación humano o no-humano mayoritaria sea la aceptada por votos, o bien, en otros casos, mediante la aplicación del mismo test varias veces en diferentes momentos<sup>8</sup> dando a entender que la atribución de inteligencia debe estar sujeta a revisión. Schweizer en *The Truly Total Turing Test* (1998) arguye que el test debería ser aplicado no a individuos sino a máquinas capaces de reproducirse y tener descendencias en su misma “especie”. El TTTT es un sucesor del Test de Turing Total propuesto por Harnad en *Other Bodies, Other Minds* en el que propone que las máquinas deberían ser capaces de realizar actos físicos para ser consideradas inteligentes, lo que se conoce como mente corporeizada<sup>9</sup>, y que la capacidad lingüística es en realidad equivalente e indicativa de habilidad motora. Resulta evidente que una máquina que no pudiese pasar el TT pudiera pasar algunas de sus extensiones, por tanto debería ser uno de los primeros objetivos.

Como crítica, el funcionalismo del que habíamos hablado al citar a Ashby, se da también en la definición de (Turing 1950), ya que identifica comportamiento inteligente como “comportamiento percibido como humano por un ser humano”, además de ser un funcionalismo subjetivo en el que la validez de la observación es dependiente de un agente o varios, según la formulación del test, i.e. no es una métrica objetiva. Además, ésta definición es limitante, pues sólo identifica comportamientos inteligentes o

<sup>7</sup><http://www.aisb.org.uk/events/loebner-prize>

<sup>8</sup>Detallados en Moor, *An analysis of the Turing Test* (1976), lo que además indica que el término “Test de Turing” ya era usado en ese año.

<sup>9</sup>Harnad lo propone como crítica al TT, aunque Turing no dice que las habilidades lingüísticas deberían ser las únicas

pensamientos con actuaciones humanas, siendo ambas identificaciones comportamiento-pensamiento y humano-inteligente problemáticas: la identificación comportamiento-pensamiento impide a la IA dar una descripción de lo que es pensamiento más allá de “es comportamiento” y la identificación humano-inteligente impide la resolución de problemas que están más allá del proceso de pensamiento humano en formas diferentes que un simple aumento en la velocidad o precisión de los mismos.

Esta crítica al funcionalismo también está implícita en el contra-argumento de la habitación china, por John Searle, uno de los más importantes. A continuación veremos ésta y otras críticas posteriores al paper de Turing.

### 3.2. Limitaciones del Test de Turing

Más allá de las objeciones que Turing anticipa en su propio paper, es interesante comprobar la polémica que éste suscita en sus contemporáneos y sucesores. Seguiremos para el estudio de la misma el comentario de (Pinar Saygin, Cicekli y Akman 2000) y (LaCurts 2011).

Al tratarse el test de Turing de una sustitución, de una pregunta demasiado amplia (*¿Puede pensar una máquina?*), por una pregunta restringida a un objetivo (*¿Puede convencer una máquina a un humano de que conversa como un humano?*) es natural que surjan maneras de cumplir éste objetivo que quizás entren en conflicto con nuestras preconcepciones sobre la primera pregunta. Esto es, en nuestro modelo de inteligencia existen falsos positivos aparentes; casos positivos en el método de testeo que no son, en apariencia, casos positivos en lo que intentamos modelar, la inteligencia; y falsos negativos, análogamente. Podemos argüir, como Purtill en *Beating the Imitation Game* (1971), que la máquina no es necesaria en ningún caso, y que realmente lo único que demuestra el Test de Turing es la inteligencia del constructor de la máquina. Creemos que ésta crítica es vacua al no existir diferencias no esenciales entre la creación de una máquina inteligente y la crianza de un niño inteligente, si la teoría sobre los indistinguibles que presentamos anteriormente se acepta.

Al tratarse de una sustitución de habilidades inteligentes por aquellas habilidades que pueden reflejarse únicamente mediante el lenguaje escrito, se entiende que el rango de inteligencia que puede medirse es menor. Gunderson, en *The imitation game* (1964), critica la validez de esta métrica de inteligencia usando un experimento mental. Imaginemos que tratamos de adivinar si el sujeto en la habitación es humano o no. Para ello disponemos de un agujero por el que meter el pie en la habitación del sujeto que se supone debe pisar para calificar como humano. Si se construye un sistema que deje



caer una piedra cada vez que alguien introduzca el pie por el agujero, dice Gunderson, podríamos identificar la piedra como humano. Esta crítica es una ejemplificación de un falso positivo, que se sostiene en que la capacidad de reconocer humanidad a partir de poder pisar un pie es limitada. Creemos que esta crítica es sensata en general, pero obvia que las capacidades derivadas del lenguaje tienen una relación más profunda con la inteligencia que pisar un pie con la cualidad de ser humano, a saber, con habilidades lingüísticas puede explicarse a otro ser que comprenda el lenguaje cómo realizar casi cualquier tarea. En todo caso, para que el TT sea una métrica completa de inteligencia debemos asegurar que esté libre de falsos positivos y falsos negativos.

Esto nos lleva a pensar en cómo atribuimos inteligencia a otros humanos. Normalmente esto pasa por asumir que uno mismo es humano e inteligente, y mediante la interacción con otros humanos atribuir también a éstos inteligencia. La mayor parte de estas interacciones discurren, después de la identificación del sujeto en cuestión como humano<sup>10</sup>, mediante la evaluación de los actos, no sólo de habla, de dicho sujeto. La parte lingüística y de comparación de ideas a través del lenguaje es ciertamente importante, aunque no es la única. Michie, en *Turing's Test and Conscious Thought* (1992) afirma que el TT no captura el pensamiento subconsciente y que las máquinas son incapaces de generar este tipo de pensamiento. Michie da como ejemplo que cualquier ser humano debería saber cuál es el plural de una serie de palabras inventadas en inglés. La crítica consiste en que es imposible que un programador considere todas las reglas del lenguaje, como las de pluralización, y las programe en una máquina. En realidad esto puede ser innecesario al programar la máquina para que aprenda de usuarios del lenguaje reales y además, no evidencia la existencia de pensamiento subconsciente en el proceso lingüístico, sino que se trata de una regla lingüística aprendida y generalizada en la mayor parte de los humanos que conocen el idioma inglés. En cualquier caso, ante el hecho de que la máquina demuestre inteligencia que el TT no captura, existen competencias clave que el TT sí captura.

Volviendo al tema de la atribución de inteligencia es interesante considerar qué ocurriría si sustituimos al evaluador del TT por una máquina. La máquina, de poseer inteligencia equivalente a la humana, debería ser capaz de identificar correcta e incorrectamente máquinas y humanos con una tasa de errores similar a la de un humano<sup>11</sup>. Podemos considerar si para la identificación correcta y la imitación de un humano

<sup>10</sup>La atribución de inteligencia humana se simplifica normalmente como atribuible por humanos al ser una cualidad esencial de éstos.

<sup>11</sup>Si la máquina tuviese menos errores podríamos concluir únicamente que tiene una inteligencia especializada en este problema, a menos que muestre que esta especialización surge de habilidades mayores que las humanas.

correcta (ambos roles del TT) debe ser necesaria la existencia de una teoría de la mente en dicha máquina, es decir, la capacidad de atribuir estados mentales a entidades que interaccionen con la misma. La respuesta a esta pregunta determinaría si la atribución de estados mentales y la identificación de actos inteligentes es un comportamiento inteligente en sí mismo que el TT original no captura. Este test fue propuesto por Watt en *Naive Psychology and the Inverted Turing Test* (1996).

En el contexto de Inteligencia Artificial esto nos lleva al si existe distinción ontológica entre la inteligencia creada artificialmente y la inteligencia natural. Es decir, si logramos crear inteligencia, sea ésta lo que sea, ¿existe aún alguna diferencia entre lo que hemos construido y la inteligencia humana? Por ello entramos en una distinción similar a la que propone John Searle (como veremos posteriormente) entre “IA débil”, significando “la IA sólo imita la acción de una mente real, pero carece de intencionalidad al no poseer estados mentales dirigidos a elementos del mundo”, e “IA fuerte”, significando “la IA es una mente real en todos sus aspectos” (Searle 1980)<sup>12</sup>. En el argumento que comentamos más arriba las consecuencias son similares, aunque la distinción se haga irrelevante. No debe entenderse este argumento como crítica a Searle: su argumento habla sobre el constructo inteligente versus el proceso inteligente natural, siendo ambos diferentes en su estructura física (la mente humana vs la habitación china), aunque no en los resultados observables, por lo que sí existe una posible distinción razonable.

En relación a esta crítica está la habitación china de Searle, que postula una habitación que recibe mensajes escritos en chino teniendo al propio Searle dentro (que no entiende el chino) equipado con un manual que le dice qué símbolos escoger para elaborar sus respuestas a los mensajes entrantes. Searle critica que dicha habitación podría ser capaz, en teoría, de superar el TT. Sin embargo, ninguna de las partes componentes entiende el lenguaje y por tanto la habitación carece de intencionalidad, que para Searle es una parte esencial de la inteligencia. Podemos argüir que el único libro que puede hacer que Searle sea capaz de elaborar respuestas con sentido para que pueda pasar el TT es un manual que enseñe chino a Searle o un manual que traduzca entre frases del chino y el inglés perfectamente o una transformación entre frases en chino y un modelo de datos de la situación que se va desarrollando durante la conversación<sup>13</sup>, ya que cualquier conversación en chino o en cualquier otro idioma no posee una estructura

<sup>12</sup>El argumento de Searle parece implicar que es necesario un entorno del que aprender y unos órganos que provean necesidades para satisfacer, para que el pensamiento esté dirigido a éstos elementos externos a la mente. En éste caso, no hay impedimento teórico en que una máquina pueda simular todos estos sistemas.

<sup>13</sup>Por ejemplo, si el mensaje incluye “Yo soy filósofo” en chino, la máquina debería operar de forma similar a: recuerda que el adversario es filósofo y trae a la memoria de trabajo hechos relativos a cualidades de los filósofos que sean relevantes en la conversación actual de acuerdo a reglas dadas.

tan simple como una tabla pregunta-respuesta, como Searle parece entender<sup>14</sup>. Evaluar el lenguaje no es trivial en todo caso, y en una conversación se establecen hechos y relaciones entre términos (véanse juegos de palabras por ejemplo) que una máquina que pase el TT debe ser capaz de identificar y desarrollar para que la comunicación sea una imitación humana fidedigna<sup>15</sup>. Por lo tanto, o bien las instrucciones incorporan un mecanismo de comprensión de palabras en chino asociadas a la respuesta, lo que hace necesario un mecanismo recursivo sobre las comunicaciones recibidas y el modelo de la situación conversacional, o bien Searle puede entender el chino. El tema de la habitación china de Searle tiene el suficiente alcance para poder desarrollar varios artículos por sí mismo. Forma parte del movimiento anti-funcionalista o anti-conductualista, en el campo psicológico, de la década de los 80 y 90.

Más allá de la limitación del test respecto a la inteligencia, el identificar actos no inteligentes como inteligentes y viceversa, cabe preguntarse si los actos inteligentes que se identifiquen sean inteligencia humana o no. La crítica de French en *Subcognition and the Limits of the Turing Test* (1990) postula un test para la capacidad de volar en una isla del mar del Norte cuyos únicos ejemplos de objetos voladores son gaviotas. El paralelismo se traza en el momento en que una máquina capaz de volar cruza su modelo de “objeto volador”, que está basado en las gaviotas y asume que la capacidad de volar de éstas debería ser indescrible de cualquier entidad voladora. La conclusión de este experimento mental es la existencia de falsos negativos en el test, en el sentido de comportamientos inteligentes *no humanos* que pasen desapercibidos al test. De nuevo nos topamos con el problema de la definición de inteligencia, que en general toma la inteligencia asumida en humanos como base. Por ejemplo, (Minsky 1985) postula que la manera en la que los humanos piensan, dividiendo y abstrayendo la realidad en objetos y conceptos simples, es comparativamente eficiente y tal vez por eso sea una muestra de *inteligencia general* y el único modelo de inteligencia<sup>16</sup>.

Otras críticas, también de French, involucran experiencias singularmente humanas como necesarias para la inteligencia, como el ser capaces de identificar hojas secas como un escondrijo útil o la evaluación de nombres de cereales como atractivos o

<sup>14</sup>Críticas similares incluyen el “saber todas las conversaciones de longitud  $x$  previamente” de Block en *Troubles with Functionalism* y *Psychologism and Behaviorism*, que afirma que una máquina podría simplemente extraer respuestas pregeneradas desde una lista. Esto obvia que construir dicha lista es materialmente imposible al ser todas las conversaciones infinitas no numerables y dependientes del contexto y de sí mismas.

<sup>15</sup>Es posible incluso que la habilidad de la máquina para simular la cultura de su adversario haga que la identificación sea más o menos precisa. Existe material desarrollado en las ciencias sociales sobre el TT, listado en (Pinar Saygin, Cicekli y Akman 2000)

<sup>16</sup>Y por tanto, afirma dicho artículo, al ser lo más probable biológicamente nos permitiría comunicarnos con vida desarrollada fuera de la tierra (‘aliens’) bajo los principios de la selección natural.

no. Se trata de una limitación del TT clara, por falsos negativos potenciales: el hecho de que si una máquina no puede empatizar o pensar en una experiencia humana no será identificada como inteligente en el TT. Las soluciones propuestas desde la IA son o bien dar a la máquina entrenamiento en vivencias humanas o bien cambiar el test para que las preguntas relativas a vivencias personales no sean relevantes. Si la primera es una solución efectiva o la segunda una limitación a la evaluación completa de inteligencia son preguntas aún abiertas. Como extensión a esta crítica agregamos la super-expresividad (*superarticulacy* en inglés) que hace referencia a la capacidad de las máquinas inteligentes de, posiblemente, superar a los humanos en algunas tareas que éstos realizan de forma intuitiva, siendo rechazados en el TT por esto mismo<sup>17</sup>: en todo caso, esto podría ayudar a crear nuevas ideas más precisas sobre la inteligencia humana, el desarrollo de nuevos test más precisos y el aumento de las propias capacidades humanas relativas a dichas tareas.

El Test de Tokyo

---

<sup>17</sup>Crítica ya anticipada por Turing, que resuelve forzando a las máquinas a “hacerse estúpidas” en el momento que esto ponga en peligro el test.



## 4 Conclusiones

---

En este trabajo hemos expuesto perspectivas históricas referentes a la creación de máquinas inteligentes en los campos de la maquinaria computacional y la algoritmia como referentes a la base de la inteligencia y la definición de inteligencia en la parte filosófica del problema. Se ha ligado el problema de la IA con el problema de los indiscernibles y la filosofía de la tecnología además de señalar relaciones entre el TT y las disciplinas académicas que tratan el tema de la inteligencia y la maquinaria computacional (i.e. psicología, algoritmia, sociología, etc.), identificando puntos de partida para el desarrollo de investigaciones en dichas disciplinas.

Se ha hecho una diferencia entre la definición de inteligencia y los comportamientos inteligentes, permitiendo evadir el platonismo de asumir que inteligencia es una entidad, por lo que se caracteriza el Test de Turing como una forma productiva de comprobar si una entidad realiza, de facto, comportamientos inteligentes.

Se han identificado las limitaciones generales del TT:

1. Hace posible identificar máquinas carentes de inteligencia general como tales (falsos positivos).
2. Hace posible que máquinas capaces de actos inteligentes pero sin habilidades conversacionales no sean considerados como inteligentes (falsos negativos). Englobamos aquí la posible existencia de inteligencias no humanas que también pasarían desapercibidas al test.
3. En ningún caso el test avanza hacia una definición más precisa de inteligencia, siendo necesaria una meta-evaluación del test para llegar a ella.

Vemos que el TT aún así nos da una definición operativa, aunque no sea usada en

los grupos que crean algoritmia inteligente orientada a tareas<sup>1</sup>, es necesaria para la identificación de inteligencia general. Uno de los objetivos del *endgame* de la IA es la identificación y caracterización de los comportamientos inteligentes y el desarrollo de tests objetivos y reproducibles de la misma. La necesidad de colaboración con la psicología, la neurociencia y el resto de las ciencias cognitivas se hace aquí evidente.

---

<sup>1</sup>La definición operativa suele basarse en métricas de *precision* y *recall* sobre conjuntos de datos estándar, para comprobar mejoras y diferencias en ajuste de funciones y clasificación.

# A Historia del Hardware Computacional

---

Se incluye aquí un resumen y comentario sobre la historia de la maquinaria de computación, cuyos puntos clave han sido extraídos de (Wikipedia 2015).

Históricamente el computador de propósito general fue diseñado por Charles Babbage en 1833, aunque no fue construido hasta el siglo XX por, primero, su hijo Prevost que construyó una parte mínima esencial del mismo que era capaz de ejecutar programas simples en 1910, y en última instancia fue imitado por el Museo de Ciencias de Londres con materiales de la época sin éxito a finales del siglo XX. Aún así se considera que el prototipo, de llegar a cumplir sus especificaciones de diseño, sería Turing-completo, sólo limitado por la precisión numérica y la memoria total de manera similar a la que cualquier computador actual lo sería. En la primera mitad del siglo XX encontramos gran cantidad de máquinas de cómputo, incluso cajeros automáticos de cobro mecánicos y electrónicos con sistemas de ayuda al cálculo de balances, pero ninguna cumple la propiedad de ser lo suficientemente general para ejecutar algoritmos recursivos.

Será en 1944 cuando se cree el Colossus Mark II, el primer computador electrónico digital Turing-completo que se mantuvo en secreto hasta la década de 1970 ya que fue planeado también para romper códigos, y usado durante la guerra fría. Éste constructo es otro ejemplo de la confluencia de ideas que se fue gestando desde el siglo XIX con la publicación de George Boole<sup>1</sup> *Las leyes del pensamiento*, que inaugura el cálculo lógico binario (o booleano), y desarrollado por Alfred Whitehead. Claude Shannon, fundador de la teoría de la información como entropía e ingeniero eléctrico, demuestra que las operaciones de la lógica booleana pueden construirse mediante la adaptación

---

<sup>1</sup>Otros ejemplos de ésta confluencia son los trabajos en lógica de Russell y Whitehead, el proyecto formalista de Hilbert o los teoremas de Gödel.



a circuitos eléctricos entendiendo el símbolo básico 0 como la ausencia de corriente eléctrica y el símbolo 1 como la presencia de la misma. La idea de usar la lógica simbólica como base de todas las matemáticas, y la idea de *Gödelización* como transformación de enunciados en números y su posterior procesado, confluyen también en la generación de la idea de la Máquina de Turing Universal ya presente en *On Computable Numbers* y consistente en la aplicación de una máquina de Turing que puede aceptar como entrada otras máquinas de Turing transformadas en números naturales y ejecutarlas. Se puede demostrar también que una máquina de Turing, para ser Turing-completa, únicamente precisa del uso de dos símbolos, 0 y 1.

Cabe notar que éste computador era programable en el sentido de que permitía la interconexión arbitraria de elementos que no es que tuviesen programas almacenados sino que eran módulos de cálculo especializados en una tarea concreta referida al cálculo booleano o de aplicación criptográfica, no porque permitiese especificar un programa a nivel de memoria como se hace actualmente, siendo gran parte de estas tareas gestionadas por el sistema operativo. Para la carga en memoria de programas se tendría que esperar a la arquitectura von Neumann y al desarrollo de los chips y memorias magnéticas.

En la década de los 50 todas estas ideas se amplían y se comercializan. La invención de la microprogramación, hoy conocida como *firmware* que permite la adición de nuevas instrucciones programadas sobre una máquina física, sin cambiar la máquina, simplifica el desarrollo de las máquinas de cómputo haciéndolas más flexibles. En 1955 los computadores sustituyen los circuitos basados en tubos de vacío por circuitos basados en transistores, reduciendo su tamaño en órdenes de magnitud y aumentando su velocidad en igual medida. Ésto se suele nombrar en la literatura como “segunda generación”.

En la década de los 60 se produce el gran salto de la computación al consumo generalizado, no por parte de la inmensa mayoría de la población sino por instituciones y grandes empresas, lanzándose el primer computador que podríamos llamar “popular”, el IBM 1401. Consistía en un computador configurable que se podía alquilar por unos veinte mil euros mensuales, al cambio actual, lo que permitió a un gran número de entidades entrar en el uso de las computadoras. Lo que hoy conocemos como mainframe requería que un ingeniero configurase el computador y cargase los programas ya escritos en el mismo y preparados para el lenguaje de la máquina que los ejecutaría, decidiendo cuánto tiempo debería gastar cada uno y en qué orden: la automatización de estas tareas dará lugar en las próximas décadas a los sistemas operativos, el uso de mainframes en tiempo compartido y en última instancia a los sistemas distribuidos y paralelos.

También de esta época son los primeros lenguajes de programación por encima del nivel de lenguaje máquina, LISP y Fortran, que requerían una etapa de traducción intermedia y aún se usan hoy en día. LISP es uno de los lenguajes clave en el desarrollo de la IA en las últimas décadas, ya que, basado en el cálculo lambda de Church, es un lenguaje que permite expresar programas de procesamiento de listas mediante listas de instrucciones donde datos y programa están expresados en el mismo lenguaje. Los primeros sistemas operativos, como OS/360 cuya historia de desarrollo está narrada en *The mythical Man-Month*, también son de ésta época. En la era del mainframe cada una de las máquinas solía ser enviada al cliente con un sistema operativo adaptado a sus necesidades específicas, los sistemas operativos de consumo general, configurables por sí mismos, no aparecerán hasta la generalización de componentes de consumo que permitan que el sistema operativo pueda ser manufacturado por separado de la máquina, e integrado en varias máquinas.

Los sistemas de tercera generación (en esencia circuitos integrados conectados mediante cables) y de cuarta generación (circuitos integrados conectados mediante circuitos integrados) serán desarrollados a mediados de la década de los 60 y de la década de los 70 respectivamente. Dado que las diferencias entre ambas generaciones sólo aparecen durante el proceso de manufactura de los mismos hemos decidido obviarlas y centrarnos en los efectos que produce sobre el comercio de computadores: mayor velocidad en menor tamaño y menor precio. Para finales de la década de los 70 el computador personal ya era un hecho y los Apple II, Commodore e IBM-PC podían verse en hogares de todo el mundo, incluida España, donde se considera que hubo una *edad de oro* del software durante los años 80.

Entre los años 80 y nuestra era los computadores han ido incrementando su capacidad computacional en velocidad y capacidad de memoria de acuerdo con la Ley de Moore, que postula que el avance de tecnología permite introducir el doble de transistores en una placa del mismo tamaño más o menos cada dos años. Ésta ley se ha topado recientemente con la limitación práctica de que el calor disipado y energía consumida por un computador altamente integrado también aumentan, precisando que dicho computador posea refrigeración y alimentación eléctrica no razonables. También, teóricamente, se postula un límite duro para dicha ley en el momento que los transistores no puedan funcionar debido a su relación de tamaño con la escala de los fenómenos cuánticos. Por estas razones hoy se tiende a aumentar la eficiencia de los sistemas no aumentando la cantidad de operaciones en serie, sino que se aumenta la cantidad de operaciones que pueden hacerse simultáneamente mediante paralelización o distribución de tareas.



## B El giro cognitivista

---

Veremos a continuación un pequeño resumen<sup>1</sup> de la segunda mitad del siglo XX que explica la convergencia de diversos grupos científicos de lingüistas, psicólogos, neurocientíficos, economistas, filósofos y computólogos en lo que hoy se conoce como *ciencias cognitivas*.

Definimos *ciencia cognitiva* como toda ciencia que se ocupa de uno o más aspectos de los fenómenos de la cognición. Como tales, las ciencias cognitivas forman un conjunto de ciencias y al mismo tiempo un campo de investigación interdisciplinar cuyo tema central es el estudio de la cognición humana, animal y mecánica (Martínez-Freire 2005, p.20), que es a la vez la fuente, se cree, de los comportamientos inteligentes. Se distinguen dos conceptos de cognición: cognición A, aquella que se refiere a la acción de tomar en cuenta una realidad o, dicho de un modo más apropiado a la terminología científica, como recepción de información y cognición B, como el uso y manejo de dicha información. Tanto en el primer y segundo caso se hacen asunciones sobre la capacidad del ser humano de recibir o hacer uso de cualquier información, así como la existencia en cierto grado de dicha información. Existen múltiples concepciones de la información que sirven de base a conocimientos tan diversos como la medición cuantitativa del contenido de un texto fuente (p.e. la concepción informativa de Shannon como desorden, que ya hemos visto en el capítulo anterior) o la fundación de la capacidad deductiva de la lógica clásica (p.e. la concepción objetivista informativa de Corcoran).

Tal y como lo expresan Allen Newell y Herbert Simon en *Human Problem Solving* (1972) el ser humano y su capacidad intelectual, y al mismo tiempo las máquinas formales de cómputo (i.e. las máquinas Turing-equivalentes) pertenecen al género, por analogía biológica, ‘sistema de procesamiento de información’. La constitución de las ciencias cognitivas como grupo de interés en las funciones cognitivas de animales, hu-

---

<sup>1</sup>Una versión extendida del comentario realizado por el autor en la asignatura F<sup>a</sup> de la Mente al nacimiento de las ciencias cognitivas.

manos y máquinas, comienza en la llamada “conferencia Dartmouth”, una conferencia de matemáticos y lógicos que pone de manifiesto el auge de una disciplina informática llamada inteligencia artificial, cuyo objetivo es la obtención de comportamiento que denominaríamos inteligente en sistemas artificiales: a saber, máquinas sentientes, capaces de actuación o razonamiento autónomo. Más allá de la cognición en animales y máquinas, se comienza a sospechar que los reinos de las plantas, los hongos e incluso los virus tienen algún tipo de cognición. Por ejemplo, se ha demostrado que las plantas segregan neurotransmisores cuando se identifica una situación de estrés (i.e. aumento en la acidez del suelo o carencia de agua) y mediante esos neurotransmisores se regula su propio crecimiento <sup>2</sup>.

Otro de los momentos importantes para este grupo de interés es el nacimiento de la psicología cognitiva, manifestado en tres hechos: la fundación en 1960 del Harvard Center for Cognitive Studies, la publicación en ese mismo año de *Plans and the Structure of Behaviour* de George A. Miller y la publicación en 1967 del primer libro de texto de esta escuela de pensamiento psicológico. Con la psicología cognitiva se recupera el mentalismo, esto es, la existencia de una vida mental interna que es al menos parcialmente ajena a la materia y se deja de lado el conductismo, y por tanto el funcionalismo, como mencionamos antes. Además, el postulado de los procesos mentales que es inherente a la psicología cognitiva, asume un cierto compromiso con el computacionalismo, la existencia de procesos que podrían ser simulados o reproducidos mediante máquinas, y acerca esta escuela a los propósitos de la inteligencia artificial.

Tal y como se dice en (Ashby 1976): “La cibernética es a la máquina real (electrónica, mecánica, neural o económica) lo que la geometría es a los objetos materiales de nuestro espacio terrestre”, en el sentido de que es una descripción simplificada, y certifica más adelante con “La cibernética es entonces indiferente al reproche de que algunas de las máquinas que estudia no están incluidas en las que encontramos entre nosotros”. La palabra *Cibernética* se refiere aquí una disciplina generalista creada en la década de 1940 y que dio origen, entre otras cosas, a lo que hoy referimos como Inteligencia Artificial en la “Conferencia Dartmouth” de 1956, instituyéndose como una de las ciencias cognitivas como veremos. Cibernética, que proviene del griego *Kyvernêtes* referido al timonel que gobierna un barco, comprendía una descripción general de todos los sistemas complejos, incluidas la vida, las máquinas, la mente humana, la economía. *Que Ashby, tal y como hemos citado, entendía la economía como si fuese una máquina procesadora de bienes y capital.* y varias disciplinas matemáticas aplicadas intentando encontrar un lenguaje que permitiese expresar las interacciones en todos ellos, es decir “la ciencia del control”

---

<sup>2</sup>Ver Ramesh y col. 2015

de todos éstos sistemas. Hoy en día “cibernética” se usa en lenguaje académico como término para agrupar diferentes campos de éstas ciencias además de en la política, la sociología y los estudios organizativos y de empresa. Parafraseando a (Pylyshyn 1970) en su comentario sobre algoritmia, la cibernética es más una cosmovisión que un campo de la ciencia aislado.

El campo de trabajo de Ashby era, en términos generales, la aplicación de los principios de la ciencia de la computación a la biología y especialmente a la neurociencia. Dejando a un lado que Ashby asume que el comportamiento inteligente es claramente mecánico, entendemos que la Inteligencia Artificial no escapa a la creación de comportamientos inteligentes fuera del ámbito humano pues, en primer lugar, tal vez no interesa la reproducción exacta y completa de todos los comportamientos a los que atribuiríamos inteligencia. Se entiende todo sistema complejo, desde la maquinaria formada por múltiples máquinas simples hasta las redes sociales pasando por los sistemas neuronales como una máquina compleja cuyas leyes son: finitas y cogentes; en el sentido de que unas no contradicen a otras en el mismo sistema, es decir, que el universo descrito no entra en contradicción interna; imitables por otros sistemas de la misma complejidad y sólo dependientes de lo observable externamente.

Es decir, en el caso de Ashby aún no se había dado el giro cognitivista respecto a los fenómenos mentales, y Ashby trata la mente como un constructo mecánico cuyas piezas son irrelevantes y de las cuales el funcionamiento observable es lo único relevante. Un funcionalismo que será, 50 años más tarde, destituido en las ciencias cognitivas por dos elementos: en la inteligencia artificial por las técnicas aplicadas a problemas reales, ya que la investigación en la explicación de la mente y el conocimiento será otorgada a los psicólogos, y en la psicología por el cognitivismo, una corriente que identifica como clave en el comportamiento inteligente a los procesos mentales no observables directamente.

Aunque inteligencia artificial y psicología cognitiva son nucleares a las ciencias cognitivas no son, evidentemente, las únicas disciplinas que se centran en el estudio de los mecanismos de consciencia e inteligencia que se dan en animales y máquinas. Ciertas disciplinas, como la sociología cognitiva, la pedagogía y la filosofía de la mente, entre otras, trabajan también con este rumbo aportando diferentes perspectivas y técnicas de trabajo. Se pone el énfasis en que las ciencias cognitivas asumen una naturalización materialista de sus postulados, i.e. no existen entidades fuera del mundo físico, por tanto la mente, el espíritu y cualquier entidad a la que pueda ser asignada la función creadora de inteligencia debe tener una manifestación física y ningún tipo de atributo no-físico. Por lo tanto, todas las disciplinas que toman parte en las construcciones de

las ciencias cognitivas si no asumen esta doctrina metafísica, por así llamarla, deben compatibilizarla con ella<sup>3</sup>.

---

<sup>3</sup>Es éste el caso de la psicología, cuyo estudio de los procesos mentales ha obviado tradicionalmente los procesos fisiológicos que los desencadenan.

# Bibliografía

---

- Aristóteles (1988). *Política*. Trad. por Manuela García Valdés. Gredos.
- Ashby, William Ross. (1976). *Introducción a la cibernética*. Trad. por Jorge Santos. *An introduction to cybernetics, 1956, Chapman & Hall Ltd., London*. Ediciones Nueva Visión SAIC, Argentina.
- Franssen, Maarten, Gert-Jan Lokhorst e Ibo van de Poel (2013). «Philosophy of Technology». En: *The Stanford Encyclopedia of Philosophy*. Ed. por Edward N. Zalta. Winter 2013. URL: <http://plato.stanford.edu/archives/win2013/entries/technology/>.
- Hilpinen, Risto (2011). «Artifact». En: *The Stanford Encyclopedia of Philosophy*. Ed. por Edward N. Zalta. Winter 2011. URL: <http://plato.stanford.edu/archives/win2011/entries/artifact/>.
- Hopcroft, John E., Rajeev Motwani y Jeffrey D. Ullman (2006). *Introduction to Automata Theory, Languages, and Computation*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc. ISBN: 0321455363.
- LaCurts, Katrina (2011). *Criticisms of the Turing Test. and Why You Should Ignore (Most of) Them*. MIT. URL: <http://people.csail.mit.edu/katrina/papers/6893.pdf>.
- Lakoff, George (1987). *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press. ISBN: 978-0-226-46803-7. URL: <http://emilkirkegaard.dk/en/wp-content/uploads/George-Lakoff-Women-Fire-and-Dangerous-Things.pdf>.
- (2012). «Explaining Embodied Cognition Results». En: *Topics in Cognitive Science* 4.4, págs. 773-785. ISSN: 1756-8765. DOI: 10.1111/j.1756-8765.2012.01222.x. URL: <http://dx.doi.org/10.1111/j.1756-8765.2012.01222.x>.



- Legg, Shane y Marcus Hutter (2007). «A Collection of Definitions of Intelligence». En: *CoRR* abs/0706.3639. URL: <http://arxiv.org/abs/0706.3639>.
- Levin, Janet (2013). «Functionalism». En: *The Stanford Encyclopedia of Philosophy*. Ed. por Edward N. Zalta. Fall 2013.
- Martínez-Freire, Pascual (2005). *La importancia del conocimiento: filosofía y ciencias cognitivas*. Thema, Universidad de Málaga.
- McCulloch, Warren S. y Walter H. Pitts (1943). «A Logical Calculus of the Ideas Immanent in Nervous Activity». En: *Bulletin of Mathematical Biophysics* 5, págs. 115-133.
- Minsky, Marvin (1985). «Communication with Alien Intelligence». En: *Byte Magazine* April 1985. URL: <http://web.media.mit.edu/~minsky/papers/AlienIntelligence.html>.
- Nilsson, Nils J. (2009). *The Quest for Artificial Intelligence: A history of ideas and achievements*. 1st. New York, NY, USA: Cambridge University Press. ISBN: 0521122937, 9780521122931. URL: <http://ai.stanford.edu/~nilsson/QAI/qai.pdf>.
- Penrose, Roger (2006). *La Nueva Mente del Emperador*. Trad. por Javier García Sanz. Barcelona, España: Random House Mondadori.
- Pinar Saygin, Ayse, Ilyas Cicekli y Varol Akman (2000). «Turing Test: 50 Years Later». English. En: *Minds and Machines* 10.4, págs. 463-518. ISSN: 0924-6495. DOI: 10.1023/A:1011288000451. URL: <http://www.cs.bilkent.edu.tr/~ilyas/PDF/minds2000.pdf>.
- Pylyshyn, Z. W., ed. (1970). *Perspectives on the computer revolution*. Englewood Cliffs, NJ: Prentice-Hall.
- Quintanilla, Miguel Ángel (2000). «Técnica y Cultura». En: *Teorema* XVII/3 Agosto. URL: [www.oei.es/salactsi/teorema03.pdf](http://www.oei.es/salactsi/teorema03.pdf).
- Ramesh, Sunita A. y col. (2015). «GABA signalling modulates plant growth by directly regulating the activity of plant-specific anion transporters». En: *Nature Communications* 6. URL: <http://www.nature.com/ncomms/2015/150729/ncomms8879/full/ncomms8879.html>.
- Russell, B. (1919). *Mysticism and Logic: And Other Essays*. Longmans, Green y Company. URL: <http://www.gutenberg.org/files/25447/25447-h/25447-h.htm>.
- Russell, Stuart y Peter Norvig (1994). *Artificial Intelligence: A Modern Approach*. EN. Prentice Hall. ISBN: 0-13-103805-2.
- Searle, John R. (1980). «Minds, brains, and programs». En: *Behavioral and Brain Sciences* 3 (03), págs. 417-424. ISSN: 1469-1825. DOI: 10.1017/S0140525X00005756.

URL: [http://journals.cambridge.org/article\\_S0140525X00005756](http://journals.cambridge.org/article_S0140525X00005756).

Turing, Alan M. (1937). «On Computable Numbers, with an Application to the Entscheidungsproblem». En: *Proceedings of the London Mathematical Society*. Vol. 42. 2. URL: [http://www.cs.virginia.edu/~robins/Turing\\_Paper\\_1936.pdf](http://www.cs.virginia.edu/~robins/Turing_Paper_1936.pdf).

— (1950). «Computing Machinery and Intelligence». En: *Mind* 59, págs. 433-460. URL: <http://cogprints.org/499/>.

Tversky, Amos y Daniel Kahneman (1981). «The Framing of Decisions and the Psychology of Choice». En: *Science* 211.4481, págs. 453-458. ISSN: 00368075, 10959203. URL: <http://www.jstor.org/stable/1685855>.

Wikipedia (2015). *History of Computing Hardware*. [Online; 05/08/2015]. URL: [https://en.wikipedia.org/w/index.php?title=History\\_of\\_computing\\_hardware&oldid=673753505](https://en.wikipedia.org/w/index.php?title=History_of_computing_hardware&oldid=673753505).

Grand Master Turing once dreamed that he was a machine. When he awoke he exclaimed:

“I don’t know whether I am Turing dreaming that I am a machine, or a machine dreaming that I am Turing!”

---

From the Tao of Computer Programming