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

UNA PERSPECTIVA HISTÓRICA

Pedro Montoto García (USC)
Enrique Alonso González (UAM)

26 de junio de 2016

«*Intelligence is* what is measured by intelligence tests.»

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

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 y la necesidad de avanzar hacia un mejor modelo de experimentos y de una definición consensuada de inteligencia. 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, Psicología Mecánica

Abstract

This work intends to study in depth the concept of intelligence as described in the *imitation game* or, as it's also known, *the Turing Test*, as shown in *Computing Machinery and Intelligence* (Turing 1950). We offer a technological and philosophical bird's eye view of the evolution that ends up with this experiment and, following that, we enumerate the pros and cons this experiment has in the case of general intelligence detection. We conclude with the experimental characterization of Turing Tests and derivatives and the neccesity of both moving towards a better experiment model and to reach a consensus on the definition of intelligence. We added, as appendices, information about secondary aspects of the evolution of computing machinery and also about the evolution of psychology in order to help to understand the context this article was developed and the development these disciplines gone through after its publication.

Keywords: History of Artificial Intelligence, Philosophy of Artificial Intelligence, Alan Turing, Cognitive Sciences, Machine Psychology

Índice general

1 Breve historia de la Inteligencia Artificial

> «I propose to consider the question, "Can machines think?"»
>
> Turing 1950

Todo desarrollo relativo a la Inteligencia Artificial comienza con una pregunta de apariencia simple, como ¿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 tocaran 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 mallorquín 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[1].

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[2], 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 autómata 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

[1] 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.

[2] Verbigracia citamos el título de la publicación de Isaac Newton *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. Éstas 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 autómata 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[3]. 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

[3]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[4], en concreto del tipo $f : \{0,1\}^n \to \{0,1\}$. 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 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 lo que suponemos la base de los comportamientos inteligentes y hemos identificado los elementos mínimos del sistema, las neuronas, estableciendo una relación clara con la disciplina de la computación y la comunicación[5], por lo que dentro de la interpretación mecanicista habitual en la época

[4]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.

[5]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.

sería suficiente estudiar sus relaciones y establecer un 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 *construír 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, per se, 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 sea el único camino que podemos tomar para la detección de Inteligencia Artificial un artefacto (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 suelan tenerlo en cuenta.

Creemos que esta introducción es necesaria para despojar las ideas sobre artefactos inteligentes del misticismo que las rodea, por tanto es necesario definir con precisión qué entendemos por "construir" (o "máquina") y por "pensar". En palabras de Turing:

> This should begin with definitions of the meaning of the terms "machine" and "think". The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll[6]. (Turing 1950, apartado 1)

[6]Una empresa americana de investigación en opinión pública creada en 1930.

Normalmente definimos artefacto (Franssen, Lokhorst y Poel 2013, apartado 2.5; Hilpinen 2011, apartado 4) como una entidad en la que algunas o todas sus propiedades preexisten en la intencionalidad de un autor, siendo necesaria la preexistencia del autor naturalmente. 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 construir sistemas técnicos o artefactos que puedan comportarse de forma inteligente, tecnología que en general se conoce también como Inteligencia Artificial.

Aristóteles, en su *Física* incluye una definición similar para 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 (Franssen, Lokhorst y Poel 2013, apartado 1.1). 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 generados mediante procesos carentes de intencionalidad, es decir, asumimos que no existe ningún creador de los entes naturales o, de existir, no podemos saberlo.

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 sólamente 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, clonada 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? Turing anticipa este problema y afirma:

> Finally, we wish to exclude from the machines men born in the usual manner. [...], for it is probably possible to rear a complete individual from a single cell of the skin (say) of a man. To do so would be a feat of biological technique deserving of the very highest praise, but we would not be inclined to regard it as a case of "constructing a thinking machine." This prompts us to abandon the requirement that every kind of technique should be permitted. (Turing 1950, apartado 3)

La razón de esta problemática es que las cualidades del artefacto que lo hacen "natural" o "artificial" no son inherentes al mismo una vez el sistema técnico para la imitación ha sido lo suficientemente perfeccionado. Este hecho será de especial relevancia cuando comentemos el *juego de la imitación*, y Turing ya anticipa ésta crítica limitando lo que pueda afirmarse como máquina inteligente construida.

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: i.e. dado el caso de que ambas entidades sean indistinguibles por sus actos inteligentes, o bien se conoce el proceso de creación de ambas entidades, o bien se conoce que ambos tienen un asentamiento material diferente, uno de los cuales ha sido identificado como natural, antes de su evaluación como inteligentes.

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, y la proposición de Turing parece demasiado arbitraria dados los avances en ingeniería genética y computación basada en soportes biológicos. La mayoría de la literatura se limita a presentar diferentes modelos matemáticos equivalentes a la *Máquina de Turing*, que

explicaremos a continuación.

Alan Turing formaliza (Turing 1937) una idea intuitiva de máquina capaz de realizar cálculos [7], que hoy conocemos generalmente como *Máquina de Turing*. Dicha máquina desprovista del aparataje formal matemático consiste en una cinta en la que se pueden escribir y sobreescribir 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: escribirá en esa misma posición un símbolo s_o (que puede ser igual a s_i), se moverá una posición adelante o atrás, y 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 sobreescribible.

Se dice que un computador u otro modelo de función computable es Turing-completo si en caso de poder ser dotado con una memoria infinita, que por motivos prácticos es imposible en computadores no teóricos, fuese capaz de calcular las mismas funciones que una máquina de Turing.

Describimos pues la convención actual del artefacto computacional como soporte físico de la Inteligencia Artificial, derivada del artículo de Turing 1950, apartados 4 y 5 que se limita a describir un "computador digital" que podría llevar a cabo las mismas tareas que un "computador humano[8]" dotado de un libro de reglas de las cuales no se puede desviar y un suministro ilimitado de material de escritura. Esencialmente, el computador humano es el responsable de leer, ejecutar y controlar el estado del *programa* escrito en el libro, usando el papel para recordar los datos que sean necesarios, lo que es equivalente a la *arquitectura de von Neumann* para computadores eléctricos. Dicha arquitectura es equivalente a una máquina de Turing, siendo sus limitaciones respecto al modelo teórico esencialmente tecnológicas (i.e. como ya hemos dicho, es imposible construír una memoria infinita como la que la máquina teórica posee, vide Turing 1950, apartado 4).

Describimos aquí el equivalente moderno de la propuesta de Turing, que esencialmente no posee capacidades adicionales salvo ser más rápido y tener una menor tasa de fallos, lo que en la jerga del ingeniero en computación sería la "infraestructura", y

[7] Que Turing usaría, probablemente sin formalizar, 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. Sobre la manera en la que el artefacto precede a su teorización en la ciencia de la computación puede leerse en (Alonso 2006, apartado 1).

[8] Antes de la aparición del computador doméstico se llamaba *computers* a las personas que realizaban cómputos, aunque en general sólo en contextos científicos.

que se compone normalmente de estos cuatro elementos:

- 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 distribuído) 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 algoritmia los más relevantes hoy en día.

Como vemos, cada parte del sistema de cómputo establece abstracciones sobre el elemento anterior dando lugar a la posibilidad de construir elementos más complejos: los algoritmos se escriben en lenguajes de programación, que se compilan o interpretan para una máquina de cómputo y un sistema operativo. La ventaja consiste en que los lenguajes de alto nivel dan facilidades para escribir algoritmos más complejos que los conjuntos de instrucciones mínimos de cada máquina particular. A su vez los sistemas operativos generalizan funcionalidades sobre maquinarias diferentes, por ejemplo, una única forma de interacción para los diferentes discos duros y sistemas de archivos. Cada una de estas partes puede considerarse como un problema 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. Sobre la evolución de las máquinas de cómputo modernas puede leerse el

apéndice A. De la relación entre estos elementos y sobre las necesidades que provocan su aparición puede leerse (Ceruzzi 1998).

Para el futuro de la computación se consideran hoy relevantes 3 tecnologías que podrían sustituir la computación electrónica: computadores fotónicos, cuánticos y bioquímicos. La expansión de la tecnología electrónica mediante elementos ópticos, con la promesa de mayor velocidad al ser las interacciones entre fotones más rápidas que las interacciones entre electrones. También la computación cuántica, que utiliza los fenómenos a nivel de partícula atómica, lo que permite la paralelización masiva de las operaciones de cálculo. Y, finalmente, la computación basada en ADN y biología molecular, basada en las interacciones químicas entre proteínas y ácidos o en los intercambios químicos entre células como en las sinapsis neuronales. 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 [9] podrían darnos esa capacidad de imitación a todos los niveles que mencionamos anteriormente.

[9]Por ejemplo, se asume que la computación basada en elementos biológicos resulta más lenta que su equivalente electrónico pero permite el procesado 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

> "Viewed narrowly, there seem to be almost as many definitions of intelligence as there were experts asked to define it."
>
> R. J. Sternberg, citado en Legg y Hutter 2007

Entender qué es inteligencia es un apartado importante aunque de difícil solución en el campo de la filosofía de la Inteligencia Artificial y de la psicología, tal y como se puede comprobar en la diversidad y número de definiciones en (Legg y Hutter 2007). En primer lugar, afirmar que existe una "inteligencia" como entidad en el mundo nos parece un platonismo peligroso, al no poder ésta observarse directamente como veníamos avanzando, y por ello hemos propuesto reemplazarla por "comportamientos inteligentes relativos a tareas para las que es necesaria inteligencia", precisamente como hace Turing. "Inteligencia", como venimos comentando, no es predicable directamente de agentes sino que, a través comportamientos que califican a dicho agente como inteligente, usamos esta palabra a modo de abreviatura que engloba todos estos actos (Diamant 2015). Turing describe la inteligencia de forma similar en su *juego de la imitación*.

3.1. *Computing Machinery and Intelligence*: El juego de la imitación

Pasamos ahora a comentar el artículo que da título a este trabajo, históricamente la fundación y guía (Hayes y Ford 1995) de la investigación rigurosa hacia máquinas con comportamientos inteligentes, *Computing Machinery and Intelligence*. Turing define por primera vez un método de detección no formal pero reproducible de lo que es un comportamiento inteligente. Aunque dicho artículo ha ido perdiendo relevancia en las ciencias de la computación a medida en que éstas se han ido acercando a la resolución

de problemas concretos en lugar de la creación de inteligencia general, sigue siendo un artículo vigente en investigación filosófica.

En resumidas cuentas, Turing enuncia las reglas del juego que permitiría descubrir si una máquina está pensando o no. Este juego consiste en:

- Se escogen 3 sujetos humanos, una mujer, un hombre y un interrogador.
- El interrogador tiene el objetivo de adivinar quién es el hombre y quién la mujer. Turing no define el sexo del interrogador, aunque se refiere a "él" suele interpretarse que es símplemente por economía de lenguaje.
- El hombre debe convencer al interrogador de que es una mujer.
- La mujer ha de ayudar al interrogador a tomar la decisión correcta.
- Los participantes no deben poder verse.
- Los canales de comunicación deben de estar anonimizados de tal forma que no den pistas de quién es quien. Se sugiere comunicación únicamente escrita.

Seguidamente pasa a considerar qué ocurriría si sustituyésemos al hombre por una máquina capaz de realizar conversaciones imitando a humanos, i.e. la máquina debe convencer al interrogador de que es una mujer humana. En todo caso, el valor del juego de la imitación es que define un experimento para la detección de un tipo de comportamientos inteligentes, la habilidad de entender y responder a las interacciones en una conversación a 3 bandas.

Hay un punto que suele escapar a las discusiones sobre este artículo sobre el que nos gustaría llamar la atención y es que Turing 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?*" (Turing 1950, apartado 1)

Es decir, el juego no trata de identificar **siempre** a la máquina como humana sino que los errores de identificación de un humano entre mujer y hombre sean estadísticamente parecidos a los que se cometan en la identificación entre mujer, hombre o máquina. Si un hombre humano tratando de imitar una mujer falla, algo que no está fuera de lo concebible, ¿deja de ser inteligente o no? Por ello se hace esta distinción estadística. Es decir, Turing acepta el error en la identificación y lo hace parte de la prueba, siendo el error en asignación de objetos a categorías inherente a la cognición humana, por ser éste un proceso subjetivo (Lakoff 1987). También, comúnmente se considera que la interpretación correcta del juego no consiste en identificar la máquina como mujer, sino como humana símplemente y sin embargo, el hecho de que al interrogador no se le indique que podría estar interrogando a una máquina tiene consecuencias tan profundas que podría considerarse un juego completamente diferente (Hayes y Ford 1995, apartado 2). Dado que la interpretación estándar considera irrelevante el sexo de los participantes (Oppy y Dowe 2016, apartado 3.1), usaremos también esta interpretación.

Turing pasa después a describir las máquinas que podrían ser programadas para participar en el test, que salvo mejoras tecnológicas se corresponden con el nivel máquina de la arquitectura de computador digital que describimos en el epígrafe 2. El resto de niveles de abstracción, como ya hemos explicado, facilitan la implementación del algoritmo proveyendo abstracciones sobre la máquina básica de cómputo pero no pueden aumentar las capacidades teóricas de la máquina. Por analogía, de la misma manera que al realizar deducciones matemáticas no partimos obligatoriamente del concepto de número: el matemático en general se apoya en abstracciones y teoremas ya demostrados. Turing indica que lo relevante en éste constructo no es la máquina en sí, sino el algoritmo que produce los comportamientos inteligentes:

> "But in view of the universality property we see that either of these questions is equivalent to this, "Let us fix our attention on one particular digital computer C. Is it true that by modifying this computer to have an adequate storage, suitably increasing its speed of action, and providing it with an appropriate programme, C can be made to play satisfactorily the part of A in the imitation game, the part of B being taken by a man?"" (Turing 1950, apartado 5)

Turing a continuación anticipa una serie de 9 objeciones, numeradas, que impedirían la construcción de dicha máquina, que reproducimos aquí agrupándolas en tres grandes categorías y actualizando las críticas a dichas objeciones (reflejamos el número en el artículo en negrita). Primera categoría, "La máquina no puede poseer cierta característica

humana":

1 Las máquinas carecen de alma.

2 Las consecuencias de la máquina inteligente son temibles ya que el hombre no debe ser sobrepasado (Interpretado literalmente por Turing como "El ser humano debe estar por encima de todas las cosas").

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.

9 Las máquinas no pueden tener Percepción Extra-Sensorial.

Estas objeciones son respondidas por Turing con contraargumentos del tipo "en realidad, no se puede saber que las máquinas no puedan tener, o incluso que los humanos tengan dicha característica". Consideramos, más formalmente, que 1, 2, 4 y 9 no afectan al test al ser éste funcionalista[10] 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" siempre que X sea una habilidad evaluable en conversaciones, una contra-crítica equivalente que es la que Turing parece esgrimir.

Segunda categoría, "Objeciones matemáticas". Al ser la base de la máquina inteligente propuesta por Turing un constructo capaz únicamente de realizar matemática sobre conjuntos discretos (Turing-completitud, recordemos), surgen las siguientes objecciones:

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).

[10]Una de las pocas alternativas a 4 es caer en el solipsismo, asumiendo que una máquina o humano que pase el test (intercambiables, recordemos) puede no ser consciente mientras que uno mismo lo es.

7 La máquina es un artefacto limitado a dominios discretos mientras que el cerebro humano opera sobre dominios continuos.

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 hay humanos que también la darían. Ésta crítica a 3 se usa también como base para hablar de la *superarticularidad*, consistente en el aumento de capacidades intelectuales humanas como resultado de la competencia con máquinas más inteligentes que éstos:

> "There would be no question of triumphing simultaneously over all machines. In short, then, there might be men cleverer than any given machine, but then again there might be other machines cleverer again, and so on." (Turing 1950, apartado 6, punto 3)

Como respuesta a 7 Turing afirma que un ajuste discreto a una función continua, digamos, calcular el valor exacto de π, podrían darse resultados suficientemente similares a los que **un humano daría** para las preguntas formuladas en el juego (i.e. números de precisión finita suficientemente cercanos al valor de π como 3,1415), y, en todo caso, como ya hemos visto las neuronas operan fundamentalmente en términos discretos.

Tercera categoría, la objeción número **8**: "El comportamiento humano es arbitrario". Ésta objeción consiste en que no existe un algoritmo o una tabla de reglas para calcular o predecir el comportamiento de un sujeto humano en todos las situaciones posibles. El contraargumento de Turing afirma que no es sensato afirmar que los humanos no "funcionan" bajo tales reglas (en cuyo caso serían ciertamente una máquina, fuere o no Turing-completa) ya que no se ha demostrado, mediante la ciencia, que no existan. Como extensión afirma que si tuviésemos un programa que responde a un número de dieciséis cifras con otro de forma arbitraria un observador podría similarmente concluir que es completamente arbitrario y carece de leyes de comportamiento, aunque sabemos que alguien ha escrito un código para este funcionamiento.

En la parte final del artículo se describe el proceso de *aprendizaje máquina*, que ha devenido en un campo propio en la inteligencia artificial. Turing denomina *mentes supercríticas*, por analogía con los materiales atómicos, aquellas en las que la inyección de material nuevo, ideas y experiencias, provoca una reacción que conduce a ideas nuevas de forma continuada[11]. Turing presupone que un método aleatoreizado puede

[11]Por la misma analogía, Turing describe las mentes animales como *mentes subcríticas*, que responderían a cada nuevo estímulo o experiencia sin generar ideas nuevas, e incapaces de generación de ideas a partir de ideas.

ser mejor para la generación de este tipo de aprendizaje, ya que no cree ver reglas en el mismo. El aprendizaje máquina plantea como hipótesis que la mente del niño es aún así sustancialmente más simple que la del adulto, siendo la mente del adulto la mente del niño añadiendo experiencias de aprendizaje y generando ideas de forma supercrítica. Por tanto, suponemos que el modelado de la mente infantil sería más simple y podríamos diseñar el programa de forma que aprendiese presentándole experiencias, de tal forma que en cuanto el proceso generase un estado del programa "adulto" pudiese pasar el Test de Turing. Creemos que es relevante como relación de un método de generación de inteligencia, pero el núcleo del artículo es la descripción del test y sus objeciones.

3.2. Limitaciones del Test de Turing

Ahora que hemos comentado el artículo de Turing estamos en posición de enumerar críticas al experimento y contraargumentarlas o aceptarlas. Existen críticas metodológicas y críticas a los fundamentos epistemológicos y psicológicos principalmente.

3.2.1. Críticas al método

Al tratarse de una sustitución de habilidades inteligentes por aquellas habilidades que pueden reflejarse únicamente mediante el lenguaje escrito, siendo los comportamientos inteligentes humanos más diversos, se puede deducir que el conjunto de capacidades de acto inteligente que pueden detectarse es menor que el conjunto total de actos inteligentes que es capaz de desarrollar un humano. Gunderson, citado en (Pinar Saygin, Cicekli y Akman 2000, apartado 3.1), usa un experimento mental análogo al Test de Turing, el *juego del pisar pies*, para demostrar que su potencial de detección de inteligencia es de hecho menor: Imaginemos que tratamos de adivinar si el sujeto en una habitación es una mujer o un hombre, para lo que disponemos de una rendija por la que meter el pie. Se supone que si recibimos un pisotón, es que hay un humano dentro y podríamos distinguir si es mujer u hombre por las "cualidades" del pisotón, pero 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. De acuerdo a Gunderson, aunque los sistemas automáticos que dan pisotones con piedras fuesen capaces de pasar este test, no constituiría prueba de humanidad en ningún caso. 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.

Un falso positivo es un resultado positivo en un test binario, de resultados posibles positivo o negativo, que debería ser negativo. Por ejemplo, en el Test de Turing la identificación de una máquina que no posee inteligencia general humana como humana. En el caso del Test de Turing surgen principalmente del hecho de que se trata de una prueba indirecta, que trata de probar algo que no se puede probar directamente ("ser inteligente") mediante un método auxiliar ("ser capaz de ser identificado como humano"). Creemos que la crítica de Gunderson es sensata en general, ya que el experimento de Turing tiene posiblemente falsos positivos, 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 Test de Turing fuese un experimento fiable deberíamos poder asegurar que esté libre de falsos positivos y falsos negativos, o al menos poder limitarlos.

Para lidiar con el problema de los falsos positivos y negativos existen reformulaciones posteriores del Test de Turing que los admiten, integrándolos en el proceso de prueba[12]. Un ejemplo es el usado en el premio Loebner (Pinar Saygin, Cicekli y Akman 2000, apartado 6.1) cuya primera edición consistió en un Test con identificación humano o no-humano votada por 10 interrogadores. Otra manera de lidiar con la posibilidad de error es añadir más pruebas continuadas, estableciendo una especie de definición inductiva de inteligencia cuyo valor vaya corrigiéndose a medida que las pruebas se van sucediendo. Moor (Pinar Saygin, Cicekli y Akman 2000, apartado 3.4) entiende que éste es el caso y propone un test continuado que se usa como *toma de datos* progresiva para tomar una decisión. Moor acepta que el Test es funcionalmente adecuado para la evaluación de todos los aspectos de la inteligencia, siendo además extremadamente complejo de pasar dado que tener una sola habilidad cognitiva compleja aprendida no es suficiente, hay que tener una serie de habilidades complejas. Sin embargo, Moor comenta que si la máquina tuviese condiciones de operación interna que obligasen a revisar la decisión tomada, sería necesario hacerlo, pero no identifica qué condiciones se requieren, por lo que consideramos ésta parte del comentario vacía de contenido. También, Block (Pinar Saygin, Cicekli y Akman 2000, apartado 4.1) considera que esta falibilidad de los humanos como un problema intrínseco al Test, pero parece argüir que un número suficientemente grande de jueces sería capaz de alcanzar un consenso sobre la inteligencia de una entidad determinada. Como sistemas más rigurosos se propone que las máquinas sean capaces de realizar, en lugar de un Test normal, un

[12]En (Pinar Saygin, Cicekli y Akman 2000, apartado 6.1) se llega a afirmar que el Test *nunca* se ha aplicado tal y como describe Turing, con 3 participantes y una sola vez.

examen general con preguntas estandardizadas y mensurables sobre sentido común y conocimiento del mundo (Clark y Etzioni 2016).

Como comentario a los test múltiples que hemos presentado, entendemos que varios jueces humanos, de ser capaces de detectar correctamente la inteligencia más de un 50 % de las veces tendrían más exito de media que un juez humano en solitario, sin embargo, si su capacidad para detectar inteligencia fuese menor que un 50 % de decisiones correctas sus errores se multiplicarían. Por lo tanto, deberíamos ser capaces de establecer algún tipo de métrica de errores para poder enmendarlos. En este caso retomamos la idea que avanzábamos en el análisis del Test, de usar la identificación correcta o incorrecta en el test sin máquinas como un *grupo de control*, para tomarlo como referencia del juicio de cada uno de los interrogadores y poder corregirlos.

3.2.2. Críticas a los fundamentos epistemológicos

Teniendo en cuenta que el Test de Turing trata sobre la detección de *otra mente*, siendo las otras mentes no observables directamente, lo cual es un problema epistemológico: tengo razones suficientes para creer que yo mismo tengo mente (o consciencia, o inteligencia), pero *¿cómo podría tenerla de otros seres humanos?*. Turing argumenta que no debería ser un problema (Turing 1950, apartado 6, punto 4), ya que considerar seriamente que no existe la mente de una entidad desconocida que podría ser un humano o una máquina es caer en el solipsismo. La razón por la que habíamos propuesto sustituir "inteligencia" por "comportamientos inteligentes" es precisamente lidiar con el hecho de que no podemos afirmar que existan entidades no observables y separar dicho problema en dos problemas más fáciles de resolver: la máquina puede realizar comportamientos inteligentes y la máquina posee inteligencia. El primero es entonces razón necesaria, aunque no suficiente, para dar soporte al segundo, por lo tanto tenemos un problema que podemos resolver aunque sus fundamentos no sean completamente correctos.

Más allá de la limitación del Test respecto a sus falsos positivos y negativos cabe preguntarse cómo han de ser los actos inteligentes que se identifiquen y qué clase de inteligencia evidencian, humana o no. Las críticas de Robert French van encaminadas en este sentido. En 1990 (Pinar Saygin, Cicekli y Akman 2000, apartado 4.5) French 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 aparece en dicha isla, algo que no encaja con su modelo de "entidad voladora" que está basado en las gaviotas y asume que la capacidad de volar de éstas debería ser indescernible de cualquier entidad voladora. La

conclusión de este experimento mental es la existencia de comportamientos inteligentes *no humanos* que pasen desapercibidos al test. French también menciona un conjunto de habilidades singularmente humanas que el Test de Turing no captura, como el ser capaces de identificar hojas secas como un escondrijo útil o la evaluación de nombres de cereales como atractivos o no, que requieren inteligencia animal en el primer caso y un entendimiento de la estética humana en el segundo. Podemos criticar el experimento de French afirmando que la manera en la que los humanos piensan[13], 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 probablemente el único modelo de inteligencia viable en un entorno evolutivo competitivo (Minsky 1985). Por tanto, afirma Minsky, al ser la cognición de tipo humano la más probable biológicamente nos permitiría incluso comunicarnos con vida desarrollada fuera de la tierra (‘aliens‘) bajo los principios de la selección natural. En cualquier caso, la crítica de French sobre que el test sólo permite detectar inteligencia humana representable verbalmente se mantiene, siendo la pregunta que necesita resolver qué elementos de la inteligencia humana son no-representables verbalmente. Millar esgrime una crítica similar (Pinar Saygin, Cicekli y Akman 2000, apartado 3.3) diciendo que la inteligencia en máquinas debe *antroporfizarse* para poder pasar el Test. En general estas críticas se deben al hecho de ser un test centrado en humanos, cuyos jueces son humanos que naturalmente tienen ciertas intuiciones sobre cómo debería actuar otro humano incluyendo cuestiones culturales, pero para ser válidas habría que demostrar que existe cognición no humana[14] o que un Test de Turing aplicado a gente de cultura diferente daría falsos negativos.

En opinión del autor de estas líneas el problema de la cultura se limita a que la habilidad de la máquina para simular la cultura de su adversario haga que la identificación sea más o menos precisa, por ejemplo si la experiencia subjetiva del juez llega a atribuír características no-humanas a alguien de cultura diferente. La influencia social sobre la máquina inteligente es ciertamente un punto interesante a explorar, ya que suele asumirse que la máquina existe sólamente dentro de la realización de un Test de Turing, aunque Turing describe un proceso de aprendizaje máquina basado en experiencias externas (Turing 1950, apartado 7) que probablemente para una identificación correcta de inteligencia serían relevantes, en el sentido de que las experiencias de aprendizaje modelan el comportamiento y la forma de expresarse de los seres inteligentes. Existe

[13] El propio French afirma la existencia de “procesos subcognitivos” que serían detectables mediante “preguntas subcognitivas”. French afirma que un Test de Turing aplicado con suficiente rigor contendría dichas preguntas implícita o explícitamente.

[14] En el apéndice B indicamos que existen indicios que podrían indicar la existencia de inteligencias no humanas en plantas y animales.

material desarrollado en las ciencias sociales sobre el Test, listado en (Pinar Saygin, Cicekli y Akman 2000, apartado 5). En todo caso, desde una perspectiva empírica, French y Miller sólo nos muestran factores que podrían hacer fallar los Test si no fuesen tenidos en cuenta, siendo preciso un sistema de corrección de los resultados obtenidos de los entrevistadores que administran los Tests.

Otras limitaciones afirman que una mente mecánica podría no tener ciertas cualidades humanas. Turing ya había respondido a críticas de éste tipo en su artículo, pero las relatamos aquí dentro de una recopilación exhaustiva. Donald Michie (Pinar Saygin, Cicekli y Akman 2000, apartado 4.3) afirma que el Test no captura el pensamiento subconsciente y que las máquinas son incapaces de generar este tipo de pensamiento, dando como ejemplo que cualquier ser humano que hablase inglés debería saber cuál es el plural de una serie de palabras, aunque fuesen 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, si la máquina es capaz de generar frases arbitrarias, al contrario que tener frases pre-generadas, entendemos que sería capaz de extraer dichas reglas del lenguaje 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 relación a esta crítica está la habitación china de Searle, (Searle 1980) y (Pinar Saygin, Cicekli y Akman 2000, apartado 4.2), 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 simbolos escoger para elaborar sus respuestas a los mensajes entrantes. Searle critica que dicha habitación podría ser capaz en teoría de superar el Test, sin embargo, ninguna de las partes componentes entiende el lenguaje y además carecen de *intencionalidad*, que para Searle es una parte esencial de la cognición humana. Por lo tanto, según Searle, podríamos detectar inteligencia con el Test pero nunca intencionalidad real. Esto está directamente relacionado con los comentarios de lady Lovelace, "The Analytical Engine (el computador Turing-completo creado por Babbage) has no pretensions to originate anything. It can do whatever we know how to order it to perform" (Oppy y Dowe 2016, apartado 2.6), que Turing ya responde en su artículo. En caso que aceptásemos esta crítica habría un test alternativo, propuesto por Bringsjord (Oppy y Dowe 2016, apartado 5.3.2), que consiste en que el test debe *meta-analizarse* para determinar si la máquina tiene de hecho intencionalidad más allá de lo programado: el método que proponen consiste en preguntar al creador de la máquina o a alguien equivalente cómo ha producido una respuesta determinada y si el

autor no puede explicarlo con referencias a las funciones básicas de la máquina, hay intencionalidad. Creemos en todo caso que el concepto de intencionalidad no está claro en ningún caso, ni siquiera para seres admitidos por consenso como vivos como los humanos.

Podemos arguír que el único libro que puede hacer que Searle sea capaz de elaborar respuestas con sentido para que pueda pasar el Test 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[15], ya que cualquier conversación en chino o en cualquier otro idioma no posee una estructura tan simple como una tabla pregunta-respuesta, como Searle parece entender[16]. Evaluar el lenguaje no es trivial en todo caso, y en una conversación se establecen hechos, y relaciones entre palabras y relaciones entre palabras y hechos que una máquina que pase el Test debe ser capaz de identificar y desarrollar para que la comunicación sea una imitación humana fidedigna.

Por tanto, o bien las instrucciones de la habitación china incorporan un mecanismo de comprensión de palabras en chino asociadas a la respuesta y un modelo del contexto relacionado a las palabras en chino, o bien Searle, concretamente, ha llegado a entender el chino en la habitación. El problema que plantea la habitación china de Searle tiene el suficiente alcance para poder desarrollar varios artículos por sí mismo y forma parte del movimiento anti-conductualista de la disciplina psicológica en la década de los 80 y 90.

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, en el sentido de que el agente sea capaz de probar soluciones que no garantizan un resultado correcto y el agente que realiza dicho comportamiento debe ser capaz de detectar esta divergencia en algún momento y corregirla, aumentando sus propias capacidades con el tiempo. Para ello el agente debe ser capaz de continuarse autónomamente durante el tiempo suficiente mientras se enfrenta a nuevos retos o crear sucesores de sí mismo que perpetúen comportamientos máquina similares. En el segundo caso la inteligencia se daría por

[15]Por ejemplo, si el mensaje incluye "Yo soy filósofo" en chino, la máquina debería recordar que el adversario es filósofo y generar o recordar hechos relativos a cualidades de los filósofos que sean relevantes. Las conversaciones en éste sentido son altamente contextuales.

[16]Críticas similares incluyen el "saber todas las conversaciones de longitud x previamente" de Ned Block (Pinar Saygin, Cicekli y Akman 2000, apartado 4.1) que afirma que una máquina podría simplemente extraer respuestas pre-generadas desde una lista. Esto obvia que construir dicha lista es materialmente imposible al ser todas las conversaciones infinitas no numerables, dependientes del contexto y de sí mismas. Otras críticas de Block se centran en criticar el conductualismo del Test, argumentando la existencia de características intrínsecas a la inteligencia que un test conductual no captura.

selección natural, dado que el entorno condicionaría la supervivencia únicamente de los individuos adecuadamente inteligentes. Consideramos, al igual que French, que el Test de Turing es suficientemente impredecible como para dar a la máquina un reto suficiente en el que comprobar que puede generar soluciones originales a problemas cognitivos, nominalmente la situación en la que una máquina debe poseer una apariencia y un carácter humanos que describir y sobre los que ser consistente narrativamente.

En este sentido social y evolutivo, Schweizer, (Pinar Saygin, Cicekli y Akman 2000, apartado 4.4.4) y (Oppy y Dowe 2016, apartado 5.3.3) arguye que el test no debería ser aplicado a individuos sino a máquinas capaces de reproducirse y tener descendencias en su misma "especie", y denomina a su versión *TRue Total Turing Test*. El TRTTT es un sucesor del *Test de Turing Total*, (Pinar Saygin, Cicekli y Akman 2000, apartado 4.4.1) y (Oppy y Dowe 2016, apartado 5.3.1), propuesto por Harnad en el que propone que las máquinas deberían ser capaces de realizar actos físicos para ser consideradas inteligentes, lo que se conoce como *mente-en-cuerpo* (*embodied mind*, en el inglés original, la traducción es mía) la idea de que ciertas categorías mentales y lingüísticas están derivadas de la información proveniente de los sentidos y la consciencia del propio cuerpo (Lakoff 2012) y que, por tanto, la capacidad lingüística depende en gran medida de la habilidad motora. Resulta difícil de creer que una máquina que no modelase correctamente este tipo de metáforas e ideas extraídas del conocimiento pudiese pasar el test de Turing, aunque en principio no parezca existir ningún inconveniente en que sean simplemente simulados sus órganos, por tanto resulta evidente que una máquina que no pudiese pasar el Test original pudiese pasar algunas de sus extensiones, por tanto hasta que se dé el caso de que una máquina pasa consistentemente el Test de Turing no debería trabajarse hacia ninguna de sus extensiones: resumiendo, bajo la hipótesis de que es posible imitar el cuerpo y la evolución de un ser vivo mediante máquinas, podemos trabajar para pasar el Test original aunque la tesis de la mente-en-cuerpo sea cierta.

Volviendo al tema de la atribución de inteligencia es interesante considerar qué ocurriría si sustituímos al evaluador del Test 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 como mínimo similar a la de un humano. 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. Podemos considerar si para la identificación correcta y la imitación de un humano correcta (ambos roles del Test) debe ser necesaria la existencia de una teoría de la mente en dicha máquina, es

decir, la capacidad de atribuír 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 Test original no captura. Este test fue propuesto, históricamente por Watt en 1996 (Pinar Saygin, Cicekli y Akman 2000, apartado 4.4.3), afirmando que es la consecuencia lógica en la comunidad psicológica de la necesidad de una definición operativa de inteligencia ante la imposibilidad de afirmar que esta existe, lo que es equivalente al problema de las *otras mentes* que mencionamos antes. Sin embargo, máquinas son utilizadas constantemente en las grandes redes sociales para detectar *spam* generado por otras máquinas, con éxito variable, y no consideramos que éstas sean particularmente capaces de inteligencia general.

Todas estas críticas convergen, en el contexto de Inteligencia Artificial, en el 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? Si aceptamos esta crítica entramos en una distinción similar a la que propone Searle: entre una cierta *IA débil* en la que la IA no es una mente humana real ya que carece de aspectos de ésta, como la intencionalidad o la consciencia, y una *IA fuerte* entendida como la IA *siendo* una mente real en todos sus aspectos. Varios argumentos que hemos presentado; vide Searle, Schweizer, Harnad y las ideas de *mente-en-cuerpo*; parecen implicar que es necesario un entorno del que aprender y unos órganos que provean necesidades para satisfacer e información de su propio estado, para que el sistema cognitivo de la máquina tenga información para desarrollarse y generar nuevas ideas. Sin embargo, no hay impedimento teórico por el momento para que una máquina pueda simular todos estos sistemas.

4 Conclusiones

En este trabajo hemos enmarcado el Test de Turing en sus orígenes filosóficos, psicológicos y tecnológicos. Hemos ligado el problema de la IA con el problema de los indescernibles y la filosofía de la tecnología, identificando un punto límite en el que el Test de Turing deja de ser relevante, i.e. la imitación indescernible de la cognición humana. Nos hemos centrado en el último capítulo en los problemas de método cientifico y fundamentos filosóficos del Test de Turing, identificando y criticando las soluciones que se proponen a los mismos.

Partiendo de las limitaciones detectadas en el Test se ha remarcado la 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, aunque tenga limitaciones claras respecto a la posible estructura de la inteligencia y la mente precisamente por su propia formulación funcionalista. En ningún caso el Test avanza hacia una definición más precisa de inteligencia, siendo necesaria una meta-evaluación del mismo para llegar a ella, caso que anticipamos al hablar de las objeciones de lady Lovelace. Es decir, se considera que para entender mejor como opera la inteligencia, de existir, debe hacerse un análisis riguroso de los comportamientos que identifican como inteligente o no-inteligente para extraer información sobre las características de los agentes que realizan dichos comportamientos. Para describir cómo podría transcurrir dicho análisis es necesario un estudio más pormenorizado de los avances psicológicos y filosóficos en la definición de inteligencia.

Se han identificado también limitaciones operativas en la aplicación de dicho Test. El Test asume que el juicio de un humano arbitrario es capaz de detectar inteligencia de una forma fiable, sin existir diferencias en tal juicio entre los casos en los que el sujeto de la prueba es un humano o una máquina. Hace posible identificar máquinas carentes

de inteligencia general como tales (falsos positivos). Siendo la prueba de inteligencia orientada a la capacidad de imitación de lo humano también hace posible que agentes capaces de actos inteligentes pero sin la habilidad de fingir ser humano, como posibles inteligencias alienígenas diferentes de la humana, animales etc., e imitadores mecánicos de éstas no sean considerados como inteligentes (falsos negativos).

Vemos que el Test de Turing aún así nos da una definición operativa, aunque ya no sea usada en los grupos que crean maquinaria inteligente orientada a tareas[17], que es necesaria por el momento para la identificación de inteligencia general. Siendo éste uno de los objetivos de la Inteligencia Artificial como disciplina, uno de los pasos intermedios debería ser la identificación y caracterización de los comportamientos inteligentes y el desarrollo de pruebas científicas que permitan detectarlos en artefactos. La necesidad de colaboración con la psicología, la neurociencia y el resto de las ciencias cognitivas para alcanzar un consenso sobre una definición completa de inteligencia se hace aquí evidente.

[17]La definición operativa suele basarse en métricas de *precision* (la proporción de respuestas correctas en el conjunto de respuestas) y *recall* (la proporción de respuestas correctas sobre el conjunto de elementos correctos posibles) sobre conjuntos de datos estandardizados en cada campo de estudio, para comprobar mejoras y diferencias entre algoritmos de clasificación, recomendación o ajuste de funciones. En realidad puede considerarse una especialización del Test de Turing que lo hace reproducible para un cierto dominio.

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 (Ceruzzi 1998).

Históricamente el computador de propósito general fue diseñado por Charles Babbage en 1833, aunque no fue construído 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[18] *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 construírse mediante la adaptación

[18]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 circuítos 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 circuítos basados en tubos de vacío por circuítos 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 distribuídos y paralelos.

Tambien 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 procesado 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 circuítos integrados conectados mediante cables) y de cuarta generación (circuítos integrados conectados mediante circuítos 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, incluída 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 una versión extendida y corregida del comentario realizado por el autor a (Martínez-Freire 2005) sobre el nacimiento de las ciencias cognitivas. Se trata de una revolución anti-conductualista durante la segunda mitad del siglo XX (Miller 2003) que resulta de la convergencia de diversos grupos científicos de lingüistas, psicólogos, neurocientíficos, economistas, filósofos y computólogos en lo que hoy se conoce como *ciencias cognitivas*. Como veremos, la interdisciplinariedad es una constante desde los años 50 del siglo XX en el estudio de la mente.

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 multidisciplinaria 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. Tal y como expresa (Newell y Simon 1972) el ser humano y su capacidad intelectual, los animales y al mismo tiempo las máquinas formales de cómputo o Turing-completas 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, humanos y máquinas, comienza en 1956 en la llamada "conferencia Dartmouth" (Miller 2003, p.142), 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 que usan para regular su crecimiento cuando se identifica una situación de estrés como un aumento

en la acidez del suelo o carencia de agua (Ramesh y col. 2015), y que las agresiones les proporcionan información sobre su entorno que usan para crecer de forma más efectiva o la necesidad de contar hasta 5 para segregar ciertas substancias (Trewavas 2016). También en 1956 Noam Chomsky publica su trabajo seminal en estructuras lingüísticas tomando como base la teoría de la información de Shannon (Miller 2003, p.142-143).

Otro de los momentos importantes para este grupo de interés es el nacimiento de la psicología cognitiva, manifestado en la fundación en 1960 de múltiples centros para el estudio y la enseñanza de este nuevo tipo de estudio psicológico (Miller 2003, p.143). Con la psicología cognitiva se recupera el mentalismo como parte de los estudios formales del comportamiento y cognición humanas, esto es, la hipótesis de existencia de procesos mentales inobservables que determinan el comportamiento observable dejando de lado el conductualismo puro, que rechazaba dichos procesos precisamente por ser inobservables. El postulado de existencia de 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. En el campo de la neurociencia la Fundación Sloan acababa de completar un programa de apoyo a dicha investigación teniendo como objetivo la unificación teórica de los sistemas neuronales con los procesos cognitivos.

En paralelo a estos desarrollos la *cibernética*, del griego *Kyvernêtès* referido al timonel de un barco, de Norbert Wiener iba ganando tracción desde sus orígenes en la década de los 40 (Miller 2003, p.142-143), una disciplina que pretendía abarcar el estudio de todos los sistemas complejos de control, incluídos la vida, las máquinas, la mente humana y la economía[19] mediante métodos matemáticos. Tal y como se dice en (Ashby 1976, p.12): "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 añade "La cibernética es entonces indiferente al reproche de que algunas de las máquinas que estudia no están incluídas en las que encontramos entre nosotros". La cibernética intentaba encontrar un lenguaje que permitiese expresar las interacciones en todos ellos, es decir "la ciencia del control" de todos éstos sistemas. Hoy en día "cibernética" se usa en lengauje académico como término para agrupar diferentes campos de estas ciencias además de en la política, la sociología y los estudios organizativos y de empresa. Parafraseando a (Pylyshyn 1970, p.103-104 y p.207-216), la cibernética es más una cosmovisión que un campo de

[19] Ashby, tal y como hemos citado, entendía la economía como si fuese una máquina procesadora de bienes y capital.

la ciencia aislado que aúna esfuerzos de forma interdisciplinaria para que los estudios sociológicos, económicos y cognitivos puedan tener el mismo rigor experimental que la física, y precede en los años 60 a la formación de las ciencias cognitivas.

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.

Bibliografía

Alonso, Enrique (2006). «De la computación a la hipercomputación». En: *Azafea, revista de filosofía* 8.8, págs. 121-146. ISSN: 0213-3563.

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.

Ceruzzi, Paul E. (1998). *A History of Modern Computing*. Cambridge, MA, USA: MIT Press. ISBN: 0-262-03255-4.

Clark, Peter y Oren Etzioni (2016). «My Computer Is an Honor Student — but How Intelligent Is It? Standardized Tests as a Measure of AI». En: *AI Magazine* 3.1.

Diamant, Emanuel (2015). «Advances in Artificial Intelligence: Are you sure, we are on the right track?» En: *CoRR* abs/1502.04791. URL: http://arxiv.org/abs/1502.04791.

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/.

Harnad, Stevan (2000). «Minds, Machines and Turing: The Indistinguishability of Indistinguishables». En: *Journal of Logic, Language and Information* 9.4, págs. 425-445.

Hayes, Patrick y Kenneth Ford (1995). «Turing Test Considered Harmful». En: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1*. IJCAI'95. Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc., págs. 972-977. ISBN: 1-55860-363-8, 978-1-558-60363-9. URL: http://dl.acm.org/citation.cfm?id=1625855.1625981.

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/.

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.

Miller, George A. (2003). «The cognitive revolution: a historical perspective». En: *Trends in Cognitive Sciences* 7.3, págs. 141-144. ISSN: 1364-6613. DOI: DOI:10.1016/S1364-6613(03)00029-9. URL: http://www.sciencedirect.com/science/article/pii/S1364661303000299.

Minsky, Marvin (1985). «Communication with Alien Intelligence». En: *Byte Magazine* April 1985. URL: http://web.media.mit.edu/~minsky/papers/AlienIntelligence.html.

Newell, Allen y Herbert A. Simon (1972). *Human Problem Solving*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc. ISBN: 0134454030.

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.

Oppy, Graham y David Dowe (2016). «The Turing Test». En: *The Stanford Encyclopedia of Philosophy*. Ed. por Edward N. Zalta. Spring 2016.

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.

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, 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.

Trewavas, Anthony (2016). «Intelligence, Cognition, and Language of Green Plants». En: *Frontiers in Psychology* 588.7. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4845027/.

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.

— (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.

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