EDICIONES COMPLUTENSE

Revista de Filosofía
ISSN: 0034-8244 • e-ISSN: 1988-284X

ARTÍCULOS

La inteligencia artificial como cuestión empírica: Un comentario de "Computing machinery and intelligence"[1]

Antonio Valor Yébenes
Universidad Complutense de Madrid

https://dx.doi.org/10.5209/resf.96771 Recibido: 29/7/2024 • Aceptado: 19/11/2024 • Publicado en línea: ?/12/2024

ES **Resumen:** En "Computing Machinery and Intelligence", de 1950, Turing acaba afirmando la posibilidad de procedimientos automáticos computacionales que sustituyan al pensamiento humano. Son reconocidas las críticas de Searle a este artículo, que se han retomado y ampliado a propósito de la actual discusión sobre el alcance y los límites de la inteligencia artificial. Especialmente relevantes al respecto resultan los últimos trabajos de Žižek y Larson, así como la defensa de las propuestas de Turing que ha realizado Daniel Dennett, especialmente en su obra final.
De toda esta discusión quiero valerme para esclarecer cuál es exactamente la posición de Turing a partir de 1950. La conclusión a la que llego es que Turing no defiende que los procedimientos automáticos computacionales, tal y como él mismo los define, sustituyen al pensamiento humano, sino que es posible que lo puedan sustituir, y que se trata de una cuestión que solo se puede resolver empíricamente.

Palabras clave: Turing; inteligencia artificial; intuición; procedimientos computacionales; algoritmos recursivos.

ENG Artificial intelligence as an empirical question: a comment on «Computing machinery and intelligence"

ENG **Abstract:** In "Computing Machinery and Intelligence", from 1950, Turing ends up affirming the possibility of automatic computational procedures that replace human thought. Searle's criticisms of this article are recognized, and they have been revisited and expanded in the context of the current discussion on the scope and limits of artificial intelligence. Particularly relevant in this regard are the latest works by Žižek and Larson, as well as Daniel Dennett's defense of Turing's proposals, especially in his final work.
I want to draw on this entire discussion to clarify exactly what Turing's position is after 1950. The conclusion I reach is that Turing does not defend the idea that automatic computational procedures, as he himself defines them, replace human thought, but that it is possible that they can replace it, and that it is a question that can only be resolved empirically.

Keywords: Turing; artificial intelligence; intuition; computational procedures; recursive algorithms.

Cómo citar: Valor Yébenes, A. "La inteligencia artificial como cuestión empírica: Un comentario de "Computing machinery and intelligence", *Revista de Filosofía*, avance en línea, 1-16. https://dx.doi.org/10.5209/resf.96771

[1] Este trabajo se enmarca dentro del grupo de investigación UCM 930174 – *Lenguaje, pensamiento y realidad*, y del proyecto de investigación PID2021-125822NB-I00 (IRENETIKA "Conflictos armados y crisis humanitarias: las Humanidades y las Ciencias sociales ante los desafíos de la seguridad multidimensional").

1. La influencia del programa formalista en Turing

El artículo de Turing de 1936, titulado "On Computable Numbers, with an Application to the *Entscheidungsproblem*", busca establecer el nexo entre los procedimientos formales algorítmicos y las máquinas automáticas, con el objetivo de superar las limitaciones del teorema de Gödel, resolver el denominado *Entscheidungsproblem*, esto es, el problema de decidir si un enunciado es demostrable o no en un sistema axiomático formal de la matemática clásica, y prescindir definitivamente de la intuición. Sin embargo, el trabajo acaba demostrando todo lo contrario, a saber, que ningún procedimiento algorítmico computable soluciona el problema de la decisión. Consecuentemente, no solo no se superan las limitaciones del teorema de Gödel, sino que se extienden de los sistemas axiomáticos formales a los sistemas computacionales. Y dada la concepción kantiana y fregeana que en ese momento se tenía de la intuición, la conclusión a la que finalmente se llega es que hay verdades que solo son aprehensibles por la intuición y el conocimiento humano, de tal manera que no es posible su sustitución por una máquina.

Gödel insistía, tras conocer el artículo de Turing, en que las consecuencias no solo se referían al ámbito de la matemática, sino a la teoría general del conocimiento, porque había saltado por los aires el empeño de Hilbert y los formalistas de prescindir de la intuición[2]. Ésta seguía siendo una capacidad exclusivamente humana cuyo uso hacía que la extensión del pensamiento fuese más amplia que la de la axiomática formal, e incluso que la de lo efectivamente computable.

Dos años más tarde, en 1938, Turing publica su tesis doctoral, que lleva por título *Systems of Logic based on Ordinals*. En el punto 11 vuelve a afirmar que el objetivo de su trabajo ha ido en todo momento en la dirección del programa formalista, intentando sustituir la intuición por una lógica formal que admita un procedimiento que haga posible la demostración de todas las proposiciones particulares. A esto es a lo que llama "ingenio". Sin embargo, reconoce de una manera tajante que, una vez más, no ha conseguido alcanzar el objetivo, porque su tesis no ofrece argumentos a favor de la eliminación de la intuición, sino del ingenio. Nos encontramos, en fin, ante un nuevo intento fallido de esquivar los resultados del teorema de Gödel y es en esto, precisamente, en lo que radica su valor.

Pasada la guerra, Turing sigue pensando en la manera de hacer compatibles las conclusiones de Gödel y las suyas propias con el ideal formalista de eliminación, o al menos reducción al máximo, de la intuición, tanto en el conocimiento matemático como en el conocimiento general. Su pretensión es, formulada en los términos que emplea en ese momento, la sustitución del pensamiento humano por máquinas computacionales. La conclusión del artículo "Computing Machinery and Intelligence", publicado en 1950, es que efectivamente tal sustitución es posible. Puesto que, con anterioridad a 1938, ya habían mostrado tanto Gödel como él mismo que hay un límite a la deducibilidad y a la computabilidad, y consiguientemente, que en cada nueva máquina computacional encontraremos límites irresolubles que requerirán del concurso de la intuición y del pensamiento humano, lo que cabe preguntarse es qué ha ocurrido entre ١٩٣٨ y ١٩٥٠ para que Turing consiga mostrar, por fin, que es posible desarrollar un sistema computacional que sustituya al pensamiento y, por supuesto, a la intuición.

2. Una nueva lectura del teorema de Gödel

En el punto 6(3) de "Computing Machinery" Turing insiste, en primer lugar, en que Gödel "demuestra que en cualquier sistema lógico lo suficientemente poderoso se pueden formular aseveraciones que no se pueden ni probar ni desaprobar dentro del sistema, a menos que el sistema en sí sea inconsistente"[3]; insiste, en segundo lugar, en que él ha descrito los sistemas lógicos en términos de máquinas computacionales digitales y éstas en términos de sistemas lógicos; y concluye que "las preguntas que no podrían ser respondidas por una máquina podrían ser respondidas satisfactoriamente por otra"[4]. Lo interesante de esta parte del trabajo es que pone a las claras que Turing ha cambiado su interpretación del teorema de Gödel. Ya no entiende que marque un límite para los sistemas axiomáticos formales y, consecuentemente, haga imposible el objetivo formalista de renunciar a la intuición, sino, por el contrario, que define una estrategia para, precisamente, superar los límites que tiene cada sistema axiomático formal en particular. Dada la relación por él establecida entre los sistemas lógicos y las máquinas computacionales, lo que finalmente acaba defendiendo es que el teorema de Gödel contiene la estrategia para construir sistemas computacionales cada vez más potentes que sean capaces de resolver aquellos problemas que no pudieron resolver los sistemas computacionales anteriores tomados como base.

La cuestión entonces es: ¿qué es lo que originó este cambio de interpretación?, ¿qué pasó entre 1938 y 1950 para que Turing en particular y una parte cada vez mayor de la comunidad de matemáticos y criptoanalistas dejasen de ver el teorema de Gödel como un límite infranqueable para el desarrollo de la

2 Copeland (2013), p. 31.
3 Turing (1950/2004), p. 450. La traducción es mía.
4 Turing (1950/2004), p. 451.

incipiente computación?[5] A mi modo de ver, son dos las razones que concurren en el transcurso de la Segunda Guerra Mundial. La primera está relacionada con el desarrollo por parte de Turing de sistemas computacionales cada vez más potentes y de máquinas más operativas y rápidas por parte de los técnicos e ingenieros que trabajaron en Bletchley Park. Es sabido que en 1939 el gobierno británico reunió en estas instalaciones a un numeroso grupo de matemáticos y criptoanalistas con el fin de descifrar los mensajes codificados que llegaban y salían de los submarinos alemanes que operaban en el Canal de la Mancha. Los cifrados alemanes se generaban en la máquina Enigma. El método ideado por Turing para descifrar consistía en eliminar un número muy grande de posibles soluciones para los códigos de Enigma buscando combinaciones en las que hubiera contradicciones[6]. Ello se hacía en una máquina de origen polaco a la que llamaban la Bomba. El número de combinaciones posibles que había que comprobar para descifrar los códigos era abrumador para la intuición o el pensamiento humano, e incluso para el trabajo colectivo que desarrollaban los cientos de personas que el gobierno utilizaba como *computadoras*. Pero, con los procedimientos ideados por Turing e implementados por un numeroso grupo de técnicos, aquella máquina podía cumplir con la tarea de simplificar posibilidades. Lo que no se podía resolver con la intuición se pudo resolver, finalmente, con el ingenio, es decir, a través de los algoritmos mecanizados.

Sin abandonar Enigma, en 1941 los alemanes comenzaron a operar una nueva máquina de cifrado llamada Tunny, con doce rotores en lugar de tres, lo que aumentaba la complejidad del descifrado. En 1942, en lo que se considera uno de los momentos más importantes del criptoanálisis y más determinantes para el transcurso de la guerra[7], Bill Tutte consigue explicar el comportamiento de Tunny. No conocía su funcionamento, sencillamente porque nunca había tenido ante él una máquina Tunny y los servicios secretos no habían conseguido ofrecerle información relevante al respecto y, sin embargo, le proporcionó a Turing un esquema de su comportamiento que permitió que, a las pocas semanas, éste ideara un método en papel para descifrar los mensajes de Tunny. La complejidad y, consiguientemente, el tiempo de descifrado aumentaba conforme los alemanes hacían más grande la red de Tunny, así que se necesitaba una máquina más potente que Bomba. Con la ayuda de Tommy Flowers, ingeniero de válvulas electrónicas y procesamientos digitales, se construyó Coloso, que fue instalado en Bletchley en 1944. En ese momento, tomando como referencia los trabajos de Turing se estaba trabajando en la Universidad de Pensilvania en la máquina ENIAC (Electronic Numerical Integrator and Computer), presentada públicamente por Von Neumann en 1946.

Lo que quiero poner de manifiesto es que Turing fue un actor principal, tanto en Gran Bretaña como en Estados Unidos a través de la influencia de sus publicaciones, del desarrollo de las máquinas computacionales, las cuales se comienzan a utilizar no solo en el ámbito militar, sino en la industria, en las finanzas, en las administraciones del estado o en las universidades. Efectivamente, hay un límite para la deducibilidad de un determinado sistema axiomático formal, como había puesto de manifiesto Gödel, y también lo hay para la computabilidad de una determinada máquina, como había mostrado Turing, pero después del desarrollo continuo de los sistemas computacionales, la cuestión es que los problemas irresolubles para una máquina tienen solución para otra máquina más potente que toma la anterior como base. De tal manera que el teorema de Gödel ya no se entiende como un límite infranqueable que tienen los sistemas axiomáticos formales y, por extensión, las máquinas computacionales, sino como una estrategia que, efectivamente, se pone en práctica para superar los límites que afronta cada máquina en particular, generando una secuencia de máquinas cada vez más potentes.

Se podría pensar que por mucho que se desarrolle esta secuencia la última máquina alcanzada tendrá un rendimiento siempre inferior a la intuición y al pensamiento, pero esto solo se sostiene sobre el supuesto, cuyo origen podemos rastrear en la filosofía de la matemática kantiana que se asume en las últimas décadas del siglo XIX y primeras del XX, de que la intuición y la deducción tienen una *ilimitada* capacidad de alcanzar la verdad. El, por así decir, *olvido* de la metafísica kantiana durante la guerra hace que este supuesto deje de operar, de tal manera que se origina una carrera entre el pensamiento humano y la máquina en la que nuevas máquinas cada vez más potentes asumen tareas hasta el momento adjudicadas en exclusiva a la intuición y al pensamiento. Se trata de una mejora continua de las máquinas que toma como estrategia el teorema de Gödel, y así las máquinas adelantan al pensamiento, que adelanta a las máquinas, que adelantan al pensamiento, que adelanta a las máquinas, sin que podamos asegurar de antemano que el pensamiento permanecerá indefinidamente en la carrera. Quizá en algún momento descubramos horrorizados los límites del intelecto humano[8], o quizá no los haya.

---

5 Un estudio detallado de esta cuestión se encuentra en Valor Yébenes (2024).

6 Turing recurrió a reglas generales, que actuaban como veloces atajos al precio de cometer errores. Las reglas funcionan cuando el número de aciertos es mucho mayor que el de errores. Esta forma de proceder es la que hoy conocemos como hipótesis de búsqueda heurística, término empleado por Simon y Newell.

7 Copeland (2013), pp. 114-115.

8 Turing (1950/2004), p. 451.

3. Funcionamientos diferentes, comportamientos semejantes

Esta relectura del teorema de Gödel, generada tanto por los trabajos técnicos en Bletchley durante la guerra como por el *olvido* de la metafísica kantiana que había inyectado Frege en la comunidad de matemáticos –precisamente para oponerse a Hilbert y al formalismo–permite al Turing de 1950 presentar la tesis que siempre había querido defender y nunca había podido, a saber, la sustitución del pensamiento humano (y la intuición) por máquinas computacionales. Hay, decía, una segunda razón que se lo permite, y es una nueva concepción de lo que es el pensamiento humano, que comienza a abrirse paso a la altura de 1945. Detrás de ello encontramos el acontecimiento técnico que supuso la construcción de Coloso para descifrar los mensajes de Tunny. Como he dicho, Bill Tutte consigue dar cuenta de su comportamiento sin tener la menor idea de su funcionamiento, gracias a lo cual Turing propone un método de descifrado. Una vez terminada la guerra, Turing y Tommy Flowers fueron enviados a Alemania con el fin de informarse sobre la tecnología que el ejército alemán había utilizado. Cuando uno de los oficiales les mostró una máquina Tunny como ejemplo de la excelencia alcanzada por la codificación alemana, Turing y Flowers sonrieron de manera cómplice. Era la primera vez que veían esta máquina y observaban su funcionamiento; sin embargo, lo sabían todo de ella, de tal manera que la habían conseguido replicar en Coloso.

Lo que este episodio pone de manifiesto es que es posible simular el comportamiento sin conocer el funcionamiento. Efectivamente, los funcionamientos de Tunny y Coloso son diferentes, pero sus comportamientos son semejantes. De la misma manera, defiende Turing en el artículo de 1950, se puede simular el pensamiento humano sin conocer su funcionamiento, en la medida en que seamos capaces de construir un sistema computacional cuyo comportamiento sea similar al de un humano. Así comienza el artículo. Si la primera lectura del teorema de Gödel, de influencia kantiana y apoyada sobre el supuesto de la *ilimitada* capacidad del conocimiento humano, nos llevaba a la conclusión de que siempre habrá comportamientos humanos que la máquina computacional no podrá realizar, desde la nueva interpretación del teorema se concluye que, dado un comportamiento humano irrealizable por una determinada máquina, cabe la posibilidad de construir una nueva máquina, tomando la anterior como base, que pueda desarrollarlo.

De esta manera argumenta Turing –¡por fin!– a favor de la posibilidad de construir máquinas que simulen el pensamiento humano, que es tanto como decir a favor de la posibilidad de explicar el pensamiento humano y sus rendimientos –especialmente la matemática– sin recurrir a la intuición ni a esas "artes misteriosas" que había supuesto la filosofía y que Hilbert y los formalistas estaban empeñados en evitar. Sin embargo, hay un argumento en contra de la posición de Turing del que él mismo se hace eco en el trabajo de 1950. Se puede decir que la máquina *simula* el comportamiento, pero no que piensa, según la definición de pensamiento desde la perspectiva en primera persona. El objetivo del presente trabajo es el de esclarecer y ponderar cabalmente la posición que acaba asumiendo Turing a partir de este momento y hasta su inesperada muerte en junio de 1954. Para ello me serviré de las muy relevantes y reconocidas críticas a "Computing Machinery and Intelligence" realizadas por Searle, las cuales han sido retomadas y ampliadas por Žižek y Larson en sus últimas publicaciones. Asimismo, tendré en cuenta la defensa de la posición de Turing en la que ha insistido Dennett en sus últimos trabajos.

4. En contra de la perspectiva en primera persona: el juego de la imitación

En el artículo de 1950 Turing propone el abandono de una definición del pensamiento desde la perspectiva en primera persona, que atribuye a modo de ejemplo a un reputado neurólogo contemporáneo suyo, Sir Geoffrey Jefferson. Según esta concepción, el pensamiento se distingue por la presencia de fenómenos psíquicos de los que carece una máquina. Pensar no es presencia del objeto sin más, sino presencia del acto psíquico dirigido al objeto. Así las cosas, una máquina puede resolver problemas o dejarlos sin resolver, pero no tiene pensamientos al respecto, porque toda la resolución que lleva a cabo no depende en ningún sentido de actos psíquicos, ni tiene conciencia de que ha hecho lo uno o lo otro, ni siente placer por lo primero, ni está enojada o deprimida por lo segundo. En definitiva, el pensamiento se define por una realidad psíquica de la que la máquina carece.

En el punto 6(4) Turing aclara que la perspectiva en primera persona no es la que se adopta cuando nos dirigimos a otra persona y nos comunicamos con ella reconociéndole pensamiento. No caemos en este caso en el solipsismo extremo, según el cual solo se puede estar seguro de que una persona piensa siendo la propia persona. Por el contrario, adoptamos con toda naturalidad una perspectiva en tercera persona y juzgamos que piensa en función de su comportamiento. Pues bien, ¿por qué tenemos que exigir a una máquina más de lo que exigimos a una persona?, ¿por qué consideramos la perspectiva en tercera persona válida para reconocer la inteligencia natural y no para reconocer la artificial? En definitiva, lo que solicita Turing es que hagamos con las máquinas lo mismo que hacemos con las personas, a saber, no caer en el solipsismo extremo y juzgarlas en función de su comportamiento. Aunque humano y máquina tengan distinto funcionamiento, como ocurría en el caso de Tunny y Coloso, si presentan el mismo

comportamiento y éste nos permite concluir que el humano piensa, también nos ha de permitir concluir que lo hace la máquina.

Esta es la razón por la que "Computing Machinery" comienza reemplazando la cuestión sobre si las máquinas piensan por el juego de la imitación[9]. Se juega por una máquina (A), una persona (B) y un interrogador (C). Este se encuentra en una habitación distinta. El objetivo es que el interrogador detecte a la máquina a partir de las respuestas ofrecidas por A y B a las preguntas que hace por escrito o tecleadas. Diremos que la máquina piensa si somos incapaces de saber si la máquina es A o B. Por supuesto, A y B tienen distinto funcionamiento, pero si teniendo el mismo comportamiento juzgamos a partir de él que la persona piensa, entonces hemos de concluir que la máquina también piensa para no discriminarla injustificadamente.

El juego de la imitación no pregunta si las actuales máquinas computacionales digitales lo hacen bien, sino si podemos pensar en máquinas de este tipo que lo hagan bien[10]. Ya sabemos que es la nueva relectura del teorema de Gödel y de los trabajos de Turing de 1936 y 1938 la que permite concluir que no hay ningún límite teórico al respecto, por lo que la cuestión es pensable. Otra cosa es que comprobemos empíricamente que hay comportamientos humanos que no son computables, en cuyo caso habría que concluir que las máquinas no piensan o, al menos, que no piensan como los humanos, y que hay una distancia insalvable entre la inteligencia natural y la artificial. Resulta entonces que la respuesta a la pregunta sobre si las máquinas piensan es empírica, y que mientras no encontremos ese límite empírico nada impide defender que las máquinas piensan. De esta manera, la carga de la prueba no se sitúa en los que afirman que las máquinas piensan, sino en los que lo niegan. El hecho hasta el día de hoy es que las máquinas computadoras digitales de estados discretos, modificadas continuamente para exhibir un almacenamiento adecuado, una velocidad cada vez mayor y un procesamiento de programas apropiado, ofrecen un desempeño cada vez mejor en el juego de la imitación[11].

5. Competencia sin comprensión

Aunque la máquina y la persona tengan el mismo comportamiento, quizá hagan algo muy distinto y, por tanto, la máquina estrictamente no piense. Turing reconoce que esta objeción es muy sólida, pero la elimina rápidamente diciendo que "si a pesar de todo, una máquina puede ser construida para jugar satisfactoriamente el juego de la imitación, no necesitamos preocuparnos por esta objeción"[12]. Efectivamente, sin hacer un análisis explícito está definiendo el pensamiento desde el comportamiento. A partir de aquí y, de nuevo, huyendo del solipsismo extremo y juzgando a las máquinas por el mismo rasero que a las personas, concluye en el punto 6(5) que "si la máquina estuviera tratando de encontrar una solución para la ecuación x2 – 40x – 11 = 0, uno se sentiría tentado a describir esta ecuación como parte del objeto del pensamiento de la máquina en ese momento"[13]. Por consiguiente, el pensamiento de la máquina, como el de los humanos, está dirigido a un objeto. Además, y en el mismo sentido, "la máquina puede, sin lugar a duda, ser el objeto de su propio pensamiento. Podría ser usada para llegar a crear sus propios programas, o para predecir el efecto de las alteraciones en su propia estructura. A través de la observación de resultados de su propia conducta, podría modificar sus programas para alcanzar algún propósito de manera más efectiva"[14]. Es decir, también la máquina se puede convertir de manera explícita en objeto de su propio pensamiento. Quizá podríamos ir más allá y afirmar que incluso de manera implícita, mientras explícitamente su pensamiento está dirigido a un objeto primario distinto de ella misma. Insiste Turing en que no está hablando de sueños utópicos, sino de posibilidades para un futuro cercano.

Esta, digamos, conciencia que la máquina tiene de sí misma, es la que le puede permitir aprender de sus errores. La idea la desarrolla Turing al final del artículo de 1950, cuando expone lo que denomina una *máquina infantil*, es decir, una máquina dotada de cámaras de televisión, altavoces, micrófonos, servomecanismos –con pocos mecanismos y muchas hojas en blanco–, de tal manera que pueda "vagar por el campo" y aprender por sí misma como el alumno que aprendió mucho de su maestro, pero mucho más por su propio trabajo[15]. Todo esto no era nuevo, porque ya en el escrito de 1948, titulado "Intelligent Machinery", había intentado una descripción de las máquinas automodificables y de los procesos educativos en máquinas[16].

Por consiguiente, tomar en serio la perspectiva en tercera persona y definir el pensamiento desde el comportamiento lleva a reconocer pensamiento en la máquina constituido por una cierta intencionalidad,

9 Turing (1950/2004), p. 441.
10 Turing (1950/2004), p. 443.
11 Turing (1950/2004), p. 448.
12 Turing (1950/2004), p. 442.
13 Turing (1950/2004), p. 454.
14 Ibid.
15 Turing (1950/2004), p. 460-461.
16 Turing (1948/2004), pp. 418-422.

en la medida en que está dirigido a un objeto, incorpora una conciencia implícita del estado de la máquina y puede, incluso, desatender al objeto e iniciar un giro reflexivo que haga explícito su estado implícito, de tal manera que la propia máquina sea capaz de modificar o educar sus estados internos. Turing ya había hablado del state of mind de la máquina computadora en el parágrafo 9 de su artículo de 1936, al afirmar que "el comportamiento de la computadora en cualquier momento está determinado por los símbolos que observa y por su estado mental en ese momento"[17].

Pero se puede argumentar en contra de la definición del pensamiento desde el comportamiento de la siguiente manera. Antes de Turing ya había miles de computadoras. Eran los hombres y mujeres que trabajaban en las distintas agencias del gobierno y departamentos científicos y de ingeniería, adiestrados para hacer cálculos de manera mecánica siguiendo repetitivamente las instrucciones que se les daban. Realizaban su trabajo sin ninguna comprensión de lo que estaban haciendo, sin ningún entendimiento de los objetivos finales ni intermedios y, por supuesto, sin ninguna creatividad. En Bletchley trabajaron cientos de computadoras humanas que hicieron posible la construcción de la máquina computadora Bomba. La máquina comenzó a realizar un trabajo de descifrado de los mensajes alemanes para el que las computadoras humanas hubieran necesitado años, pero tanto aquella como éstas estaban haciendo lo mismo, simplemente seguir una tabla de instrucciones de manera mecánica. Si afirmamos que las computadoras humanas no piensan por el simple hecho de que están siguiendo mecánicamente instrucciones, tendremos que reconocer que tampoco lo hace la máquina computadora, la cual opera de la misma manera, a la que Daniel Dennett denomina *competencia sin comprensión*[18]. En definitiva, se está realizando una acción *ciega*, que no responde a objetivos ni a relaciones de las instrucciones entre sí, y cuando esto ocurre no se puede hablar con rigor de pensamiento. Para hacerlo habría que reconocer un comportamiento con *sentido*, esto es, *competencia con comprensión*.

Una variación de este argumento es el famoso experimento de la habitación china, propuesto por Searle precisamente para rebatir a Turing[19]. Una persona que no sabe chino puede responder a los mensajes que le llegan escritos en chino utilizando manuales, gramáticas y diccionarios, es decir, aplicando reglas, de la misma manera que lo puede hacer una máquina. Pero, aunque desde la perspectiva en tercera persona reconozcamos tanto a la persona como a la máquina la *competencia* de responder en chino, esto no nos puede llevar a la conclusión de que *comprenden* el chino. Ambas tienen competencia sin comprensión, esto es, no entienden el *sentido* de las respuestas, de tal manera que la semántica del lenguaje está completamente ausente. Todo ello es lo que lleva a Searle a concluir que estrictamente no hay pensamiento.

Como indicó Turing en su artículo de 1936, en una máquina computadora quedan especificadas las operaciones de una manera completamente formal, en concreto, en secuencias de ceros y unos impresos en una cinta infinita dividida en celdas. Una *regla* de la máquina determina que cuando está en un estado y el cabezal lee un cierto símbolo en la cinta, entonces escribe el mismo símbolo o lo cambia, se mueve a derecha o a izquierda y cambia de estado. Searle insiste en que "los símbolos no tienen ningún significado, no tienen ningún contenido semántico, no se refieren a nada"[20]. Y así es, porque incluso los ceros y los unos empleados por Turing no se utilizan como números, sino meramente como símbolos vacíos de contenido de un alfabeto finito.

6. Sintaxis y semántica

Lo que Searle defiende es que tener pensamiento es más que tener procesos formales o sintácticos. "Esto es, incluso si mis pensamientos se me presentan en cadenas de símbolos tiene que haber más que las cadenas abstractas, puesto que las cadenas por sí mismas no pueden tener significado alguno. Si mis pensamientos han de ser *sobre* algo, entonces las cadenas tienen que tener un *significado* que hace que sean los pensamientos sobre esas cosas. En una palabra, la mente tiene más que una sintaxis, tiene una semántica. La razón por la que un programa de computador no puede jamás ser una mente es simplemente que un programa de computador es solamente sintáctico, y las mentes son más que sintácticas. Las mentes son semánticas, en el sentido de que tienen algo más que una estructura formal: tienen un contenido"[21].

¿Qué es la semántica, el significado, el sentido? Searle toma como punto de partida el hecho "puro y simple" de que hay estados subjetivos, eventos mentales, de los que tenemos consciencia y que son tan reales y tan irreductibles como cualquier cosa del universo[22], y que es la *intencionalidad* el rasgo que propiamente los define, lo cual quiere decir que los estados mentales están dirigidos, o referidos, o son

17 Turing (1937), p. 250.
18 Dennett (2017), pp. 62, 94-100.
19 Searle (1994), pp. 37-39.
20 Searle (1994), p. 36.
21 Searle (1994), p. 37.
22 Searle (1994), p. 19-20.

sobre, un mundo distinto de la mente[23]. Puestas así las cosas, el *sentido* tiene que ver con el contenido intencional del acto mental, mientras que el *significado* solo existe donde hay contenido intencional unido a la forma de su exteriorización. Por eso dice Searle que, estrictamente, el término "significado" solo se aplica a nociones y a actos de habla. Por ejemplo, podemos preguntar lo que significa una emisión u oración como "*Es regnet*", pero no lo que significa mi creencia de que está lloviendo[24]. Las creencias o los deseos tienen sentido, pero no significado.

Según Searle, cuando el hablante desarrolla la compleja acción de producción de marcas o sonidos que son más que marcas o sonidos porque tienen la intención de emitir un mensaje a un oyente –dicho de otra manera: cuando se realiza un *acto de habla*– hay un doble nivel de intencionalidad que se corresponde con la distinción entre *representación* y *comunicación*. Por un lado, el acto contiene la intención de representar algún hecho o estado de cosas y, por otro, la intención de comunicar, la cual se adhiere a la primera. La representación es previa a la comunicación e independiente de ella; efectivamente, puede haber representación sin comunicación, pero no puede haber comunicación sin representación. En rigor, el *sentido* está contenido en el acto de representación, y el *significado* está contenido en el acto de habla, que incorpora tanto la intención de representar como la de comunicar[25]. Con arreglo a lo dicho, puede haber sentido sin significado, pero no significado sin sentido.

En definitiva, lo que trata de defender Searle con el argumento de la habitación china es que una máquina computadora ejecuta una secuencia de acciones que son estrictamente mecánicas, esto es, que están producidas exclusivamente por reglas. En cambio, las acciones humanas no están producidas por reglas –a no ser que se nos exija actuar como máquinas computadoras–, sino por el pensamiento, constituido por una sucesión de estados subjetivos conscientes, reales e irreductibles, que se caracterizan por ser intencionales, lo cual quiere decir no solo que están dirigidos a un mundo distinto de ellos mismos, sino también que acarrean un contenido intencional de sentido y significado. En fin, los humanos tenemos *competencia con comprensión* (con sentido y significado) y las máquinas tienen *competencia* (incluso la misma que los humanos) *sin comprensión*.

Sin embargo, aunque Searle insista en ello, su argumentación no rebate el núcleo del argumento de Turing, que exige determinar si una máquina piensa con los mismos criterios que determinamos que una persona piensa, a saber, sin recurrir a la perspectiva en primera persona, sino atendiendo a la perspectiva en tercera persona, es decir, al comportamiento desarrollado por el hombre o la máquina. En cambio, el punto de partida de Searle es precisamente el de la primera persona, que es desde donde se reconoce –dice con mucha soltura y poca argumentación– la existencia de "cosas mentales, tales como nuestros pensamientos y sensaciones", "eventos mentales puros y simples", "estados subjetivos *intrínsecos* tan reales y tan irreductibles como cualquier cosa del universo"[26], de los que somos conscientes y están caracterizados por la intencionalidad.

Reconoce que las que califica como "concepciones materialistas de la mente, actualmente en boga –tales como el conductismo, el funcionalismo y el fisicalismo– terminan negando explícita o implícitamente que haya cosas tales como las mentes del modo en que las pensamos ordinariamente"[27], y afirma que responden a una "concepción «científica» del mundo como compuesto de cosas materiales"[28] que no entiende la subjetividad de los estados mentales, una subjetividad que "está marcada por hechos tales como que yo puedo sentir mis dolores y tú no puedes. Yo veo el mundo desde mi punto de vista, tú lo ves desde tu punto de vista. Yo soy consciente de mí mismo y de mis estados mentales internos, como algo completamente distinto de los yoes y los estados mentales de otras personas"[29]. Argumenta que a partir de la concepción científica surgida en el siglo XVII se ha entendido que la realidad es algo que tiene que ser igualmente accesible a todos los observadores competentes, es decir, que tiene que ser objetiva, y que esta concepción científica es la que no ha permitido reconocer la realidad de los fenómenos mentales, precisamente porque son *subjetivos*[30].

Pero, insisto, la cuestión para Turing no tiene que ver con el reconocimiento de la realidad de los fenómenos subjetivos. Aun reconociéndola, el caso es que no la tomamos como criterio para determinar que los hombres piensan, dado que solo la puedo reconocer en mí y nunca en los otros. El criterio que tomamos es el del comportamiento, y éste mismo es el que debiéramos tomar para determinar si una máquina piensa. Diremos que una máquina piensa si se comporta igual que un hombre, y que no piensa –o que no lo hace como un hombre– si en algún momento encontramos un comportamiento humano no computable, lo cual, reinterpretado el teorema de Gödel una vez arrumbada la metafísica kantiana, no se puede establecer *a priori*, sino siempre a la luz de la experiencia.

---

23 Searle (1994), p. 20.
24 Searle (1992), p. 42.
25 Searle (1992), pp. 172-174.
26 Searle (1994), pp. 19-21.
27 Searle (1994), p. 19.
28 Ibid.
29 Searle (1994), p. 21.
30 Ibid.

7. Nuevos argumentos contra las máquinas pensantes

El argumento de la habitación china ha sido ampliamente discutido hasta la actualidad[31] y reformulado mediante ingeniosas modificaciones. Me refiero ahora con brevedad a las nuevas argumentaciones que recientemente tanto Žižek como Larson han tratado de ofrecer contra la posibilidad de que las máquinas de Turing piensen.

7.1. El cambio de sentido

Žižek defiende que las acciones tienen unidad constitutiva desde un sentido que no es un contenido ingrediente de la propia acción, de tal manera que aun manteniendo la acción se puede modificar la unidad constitutiva simplemente cambiando el sentido. El ejemplo que pone es el siguiente[32]: "Estaban haciendo queso de la manera habitual, pero entonces el queso se pudrió y comenzó a oler mal, y descubrieron que aquella monstruosidad (según los criterios al uso) tenía su propio encanto; estaban elaborando vino de la manera habitual, cuando algo falló durante la fermentación, y entonces empezaron a producir champán ..." Es decir: se siguen los pasos habituales, la acción se mantiene, pero cambia el sentido y, consiguientemente, la unidad objetiva. Se desarrolla una determinada competencia –la de hacer queso o vino–, algo ocurre y el resultado obtenido se considera un fracaso. Sin embargo, el hombre, y no la máquina, operando de la misma manera y, en consecuencia, obteniendo el mismo resultado, tiene la capacidad de cambiar el sentido y generar, no a través de la acción, sino del cambio de sentido, una nueva unidad objetiva. El queso y el vino que antes se tiraban porque olían mal y sabían peor, comienzan a ser considerados exquisiteces por la aristocracia y poco después se convierten en un lujo al alcance de una minoría. Y las mismas competencias y circunstancias que antes se consideraban erróneas porque conducían al fracaso, ahora se consideran precisas, refinadas y solo al alcance de los expertos que saben ejecutarlas hasta la obtención de un resultado excelente.

Otro ejemplo que pone es el denominado *la paradoja de Hugh Grant*[33], recordando la famosa película *Cuatro bodas y un funeral*. El protagonista, Charles, un apuesto y elegante inglés, intenta expresar su amor a Carrie, una atractiva estadounidense. En primera instancia, los torpes y confusos intentos de Charles no son entendidos por Carrie como una muestra de interés. Sin embargo, a partir de un cierto momento Carrie es capaz de invertir el sentido de lo que ve y darse cuenta de la autenticidad del amor de Charles. La conducta de Charles sigue siendo la misma, pero finalmente Carrie la inserta en un excedente de sentido que modifica por completo la totalidad de su experiencia desde el comienzo de la relación. La conclusión de Žižek es que una máquina no puede captar este excedente de sentido.

7.2. La posición del yo

El segundo argumento de Žižek tiene que ver con el momento constituyente del yo, del que carecen las máquinas. Los hombres vivimos en medio de una determinada situación biológica, social, cultural, lingüística, etc. Nuestro comportamiento responde a sus exigencias, de tal manera que cada actividad prepara el camino de la siguiente. El yo surge cuando, por algún motivo, esta secuencia se rompe y la representación implícita que el sujeto tiene de sí mismo fracasa. Dice Žižek que en ese momento el sujeto aparece como yo más allá de sus condicionantes materiales, sociales, culturales, etc., sin que se trate de una nueva creación, sino de un yo que estaba ahí desde siempre. Criticando tanto el esencialismo cartesiano como el idealismo concluye que el sujeto *no es más* que esa reflexibilidad[34].

La identidad sexual es un buen ejemplo para entender la constitución del yo. Con arreglo a la situación en la que vivimos comenzamos a desplegar una determinada conducta sexual. Si ésta fracasa, el sujeto puede cambiar su identidad sexual, proyectándola sobre el pasado y el futuro y definiendo un nuevo yo desde la negatividad que se quiere superar. En este momento el yo se aprehende como auténtica *causa sui*, antes dejándose llevar por los intereses de la vida cotidiana, pero ahora, a resultas de un fracaso, absolutamente reconstituido. Recordando a Hegel, Žižek da a este momento el nombre de *contragolpe absoluto*. A partir de aquí se pregunta si una máquina puede tener yo, identidad personal, constituida a partir del enfrentamiento con una negatividad y por la fuerza del contragolpe absoluto. La respuesta es que, ante el fracaso, la máquina se paraliza y es incapaz de superarlo autoidentificándose de una nueva manera.

7.3. La libertad y el juicio moral

Como Descartes, Žižek insiste en que la identidad constituida es libre porque tiene la capacidad de aceptar o no las causas de la situación en la que se encuentra en la vida cotidiana, las cuales ejercen su poder

31 Véase al respecto Cole (2020).
32 Žižek (2023), p. 54.
33 Žižek (2023), p. 56.
34 Žižek (2023), p. 96.

solo si ha habido previamente un acto de asunción por parte del yo. Si tal acto no se produce, las causas decaen y quedan sin efecto alguno[35]. Esta es la clave de la libertad en la Segunda Meditación cartesiana que Kant desarrolla posteriormente: las razones ejercen su poder solo porque las acepto como razones.

A partir de aquí es fácil entender la responsabilidad sobre los actos propios y el juicio moral. Aunque el sufrimiento generado por mis acciones pueda ser explicado como consecuencia de las causas o razones de la situación en la que nos encontramos unos y otros en la vida cotidiana, no está justificado en la propia situación porque ésta no es inevitable. Y no lo es porque, sabiendo que las razones operan con necesidad solo cuando son asumidas, puedo no asumirlas, generando de este modo una nueva situación y una nueva cadena causal. Así las cosas, es fácil entender que una máquina no tiene libertad ni, consecuentemente, puede ser un sujeto moral.

7.4. El proyecto

Como hemos visto, el enfrentamiento con una negatividad en el seno del contragolpe absoluto hace posible la suspensión de causas y razones que determinan nuestra situación en la vida cotidiana y la posición del yo libre. Pero añade Žižek que, por sí solo, este momento no permite la superación del fracaso porque aún no ha quedado constituida una nueva situación para el yo. Para ello se necesita la propuesta de un proyecto superador que oriente la acción futura en el seno de la nueva situación.

Žižek insiste en que es el yo libre, liberado de su situación, el que se da el nuevo proyecto con miras a la superación del fracaso y la constitución de una nueva situación. El yo se da el proyecto y comienza a vivir en el *tiempo del proyecto*. Ejercida la libertad y hecha la elección, la acción futura se despliega con arreglo al proyecto. En este punto la diferencia entre el hombre y la inteligencia artificial es brutal: la máquina, el software, Neuralink, no ponen el proyecto, sino que son el resultado de un proyecto que no han puesto.

7.5. La abducción

Larson argumenta a favor de la especificidad del pensamiento humano defendiendo que éste es capaz de realizar un tipo de inferencia que no puede realizar una máquina de Turing: se trata de la abducción. Las máquinas computadoras infieren por deducción o por inducción, pero no es posible que lo hagan por abducción[36].

Ch. S. Peirce entiende la abducción como "el proceso de formación de una hipótesis explicativa. Es la única operación lógica que introduce una nueva idea"[37]. "La abducción consiste en estudiar hechos e inventar una teoría que los explique"[38]. La forma del argumento abductivo es, según Peirce, la siguiente[39]:

- El hecho sorprendente, C, es observado;
- pero si A fuera verdadera, por supuesto se daría C;
- luego hay razón para sospechar que A es verdadera.

Por medio de la abducción "se selecciona tentativamente como la más *razonable* aquella hipótesis, de entre las que compiten entre sí, que, a criterio del investigador, mejor compatibilidad muestra con los datos disponibles"[40]. Efectivamente, surge un nuevo enunciado general del cual se pueden deducir los enunciados particulares que se refieren a las observaciones concretas, unas ya realizadas y otras por realizar. Y esto tanto en el conocimiento ordinario como en el conocimiento científico. La cuestión más relevante aquí es cómo surge en el conocimiento humano ese enunciado general. Frecuentemente se alude de una forma poco precisa a aspectos "instintivos" o "intuitivos". Peirce habla del *insight*[41], de un *destello* con el que se ilumina una nueva idea que se toma como hipótesis.

Una vez realizada la abducción y explicitada la hipótesis, cambian de sentido los hechos, que quedan *iluminados* bajo la luz que aporta la nueva hipótesis. Por ejemplo, los griegos entienden que las piedras caen o los cuerpos celestes se mueven con arreglo a su propia *physis*, pero, explicitada la ley de gravitación universal, entendemos que es la fuerza de gravedad la que genera esos movimientos. Los griegos *ven* los planetas moviéndose conforme a su naturaleza y nosotros los *vemos* moviéndose por la ley de gravitación universal.

Larson insiste en que tanto la abducción como el cambio de sentido que acarrea no se realizan solo en el conocimiento científico, sino también en el conocimiento ordinario, e implícitamente operan dando

35 Žižek (2023), p. 195.
36 Una buena descripción de los procesos abductivos y del papel de la abducción en la ciencia se encuentra en Rivadulla (2010; 2015).
37 Peirce (1965), 5.171.
38 Peirce (1965), 5.145.
39 Peirce (1965), 5.189.
40 Rivadulla (2010), p. 122-123.
41 Peirce (1965), 7.202-7.207.

sentido a la comunicación. Por ejemplo, si me quiero comprar un coche y estoy con mi hijo visitando un concesionario, al decir "lo quiero" el oyente entenderá que me refiero a un determinado coche, no a mi hijo. Pero para llegar a entender eso el oyente ha tenido que abducir, atendiendo al contexto, que me quiero comprar un coche, y esta abducción es la que da sentido a los hechos observados.

Larson concluye que es precisamente la abducción lo que caracteriza el pensamiento humano[42]. Las máquinas computadoras pueden hacer deducciones, como ha mostrado la computación desarrollada hasta los años 90, e inducciones, como pone de manifiesto la actual computación desarrollada sobre internet y *big data*, pero, a juicio de Larson, carecen de ese momento de *insight* que hace posible la abducción y la iluminación significativa de los hechos que puede llevar, incluso, a su completa resignificación.

8. Una base biológica para el pensamiento, pero no algorítmica

Lo que tienen en común estos argumentos es que, como en el caso de la habitación china, recurren a la perspectiva en primera persona. Efectivamente, se hace una evaluación del cambio de sentido, de la posición del yo, del proyecto y de la abducción desde la reflexión del sujeto sobre su propia subjetividad. Se trata de la misma reflexión que, en el caso de Searle, hace posible el reconocimiento de las "realidades mentales", así como una descripción fenomenológico-descriptiva de sus rasgos "intrínsecos", como son la conciencia, la intencionalidad, la subjetividad, el sentido y el significado. Pero la crítica de Turing se mantiene: ¿si la perspectiva en primera persona no es la que utilizamos para juzgar que una persona piensa, por qué la utilizamos para juzgar que una máquina piensa?

Tanto Searle como, más recientemente, Žižek y Larson –azorados por el brutal desarrollo de la inteligencia artificial y Neuralink– siguen obviando el argumento de Turing, identificando el pensamiento con esa "realidad irreductible" formada, en concreto, por *mis* pensamientos, *mis* sensaciones, *mis* dolores, *mis* cosquilleos, *mis* estados mentales internos[43], y atribuyendo una realidad así a los hombres, pero no a las máquinas. La pregunta que me hago, referida por el propio Turing[44], es cómo se puede reconocer esa realidad en otro hombre sin ser ese hombre y negársela a una máquina sin ser esa máquina, cómo afirmar que otros también piensan, pero no las máquinas, si no podemos salir de nosotros mismos, es decir, si no podemos abandonar el solipsismo más absoluto[45].

La respuesta que doy es que lo consiguen haciendo depender los fenómenos mentales y el pensamiento del comportamiento de una realidad biológica humana, especialmente del cerebro, de su sistema neuronal y modular y del sistema nervioso central, que consideran irreductible –quizá no en parte, pero sí en su totalidad– a procesos algorítmicos y formales. El pensamiento es una realidad propia generada por la totalidad de un complejo sistema biológico e insisten en que no es, o no solo es, formal. "El cerebro no es, o al menos no es solamente, un computador digital... Los cerebros son motores biológicos; su biología importa"[46]. Consecuentemente, niegan que cualquier sistema material, por ejemplo, un computador hecho con viejas latas de cerveza que recibe energía por molinillos de viento, sea capaz de contener fenómenos subjetivos solo por el hecho de que se comporte con arreglo a un determinado algoritmo o programa computacional[47]; niegan que un número indefinidamente extenso de diferentes géneros de *hardware* sea capaz de generar fenómenos mentales y pensamiento solo por el hecho de contener un determinado *software*. Así las cosas, puedo suponer en otros hombres la realidad mental y el pensamiento que soy capaz de reconocer en mí –desde la perspectiva en primera persona– no porque pueda ser esos otros hombres, sino porque reconozco en ellos –desde la perspectiva en tercera persona– la misma base biológica que tengo yo. En definitiva, si afirmamos que los hombres piensan y las máquinas no, es porque reconocemos en ellos una realidad biológica, la nuestra, responsable del pensamiento, que no reconocemos en las máquinas ni podremos reconocer porque es irreductible a procesos computacionales.

Searle se refiere a esta dependencia entre el sistema biológico humano y el pensamiento en numerosos textos:

> Los dolores y otros fenómenos mentales son sólo rasgos del cerebro (y quizá del resto del sistema nervioso central).
>
> ...
>
> Lo mismo que la liquidez del agua es causada por la conducta de elementos del micronivel y, con todo, es al mismo tiempo un rasgo realizado en el sistema de microelementos, así exactamente en ese sentido de "causado

42 Larson (2022), p. 199.
43 Searle (1994), p. 21.
44 Turing, (1950/2004), p. 452.
45 Ibid.
46 Searle (1994), p. 47.
47 Searle (1994), p. 34.

por" y "realizado en", los fenómenos mentales son causados por procesos que tienen lugar en el cerebro en el nivel neuronal o modular, y al mismo tiempo se realizan en el sistema mismo que consta de neuronas[48].

También Žižek reafirma esta dependencia en su última obra: "La base material de este bucle de la posición del yo sigue siendo, naturalmente, ésta: no hay espíritu sin materia, si destruimos el cuerpo, el espíritu desaparece. Sin embargo, la posición del yo no es sólo una especie de «ilusión del usuario»; tiene realidad propia, con efectos reales"[49]. Por otro lado, Larson defiende el papel de la neurociencia en el estudio del pensamiento, pero lamenta que en la actualidad haya quedado encerrada en los límites de la mitología del *big data* y la informática, descuidando la importancia que el descubrimiento y la experimentación tienen en la investigación científica[50].

Tal dependencia, digámoslo una vez más, es la estrategia que les permite afirmar que no solo yo pienso, sino que todos los hombres, por el hecho de tener mi misma base biológica, piensan. Además dan otro paso, que consiste en defender que el comportamiento de la base biológica humana que genera el pensamiento no es computable en su totalidad, aunque quizá pueda serlo en parte. Este segundo paso es la estrategia para concluir que las máquinas computacionales, precisamente por no ser más que computacionales, no pueden pensar.

Pero reconsideremos este segundo paso: ¿cómo defienden que el comportamiento de la base biológica humana que genera el pensamiento no es computable en su totalidad, aunque pueda serlo en parte? La defensa se realiza mediante la argumentación que ya hemos revisado. En definitiva, reconociendo que yo y todos los organismos biológicos como yo somos capaces de generar una "realidad mental" con unos rasgos "intrínsecos" –de naturaleza *semántica*– que no se pueden generar con meras formas o algoritmos –de naturaleza *sintáctica*–. "La sintaxis sola no es suficiente para la semántica y los computadores digitales en tanto que son computadores tienen, por definición, solamente sintaxis"[51].

Con ello se nos otorga a los humanos un doble privilegio: por un lado, el de contener una subjetividad de la que carecen las máquinas; por otro, el de estar constituidos por un complejo motor biológico –del que también carecen las máquinas– que precisamente hace posible que surja esa subjetividad. Además, queda justificado que la perspectiva en primera persona sea la única válida para conocer mi subjetividad, la subjetividad humana en general y la ausencia de subjetividad en las máquinas. Puestas así las cosas, aunque la máquina supere continuamente el juego de la imitación y sea capaz incluso tanto de describir sus propios estados mentales como de caracterizarlos con rasgos "intrínsecos" propiamente humanos[52], Searle y sus sucesores siempre podrán decir lo que de hecho acaban diciendo: que el comportamiento de la máquina es una simple simulación que responde a un *software* y no a una realidad mental, o dicho de otra manera, que la máquina tiene competencia sin comprensión, incluso aunque llegue a ser capaz en algún momento presente o futuro de decir de sí misma que tiene comprensión.

9. Una petición de principio

A mi modo de ver, la argumentación empleada para criticar el juego de la imitación y negar pensamiento a las máquinas más parece una petición de principio. Y ello por varias razones. En primer lugar, porque supone la existencia de una "realidad mental" que se puede explicitar y describir fenomenológicamente en primera persona. Una realidad mental descubierta a través de una reflexión que es capaz de realizar la subjetividad sobre sí misma, en virtud de la cual no *pone* ni *quita* nada, sino que simplemente se limita a explicitar lo implícito y a caracterizarlo descriptivamente. Esta forma de proceder es propia de la tradición cartesiana y fenomenológica. El problema es que da por supuesto que la reflexión no es *objetivante* en un doble sentido: 1) no convierte al sujeto en objeto, lo que haría imposible el conocimiento del sujeto mismo; y 2) no conlleva supuestos implícitos (provenientes de las ciencias físicas o matemáticas, de la psicología, de la filosofía, del conocimiento ordinario) que distorsionan el conocimiento que la subjetividad obtiene de sí misma. El problema de la posibilidad de una *reflexión no objetivante* ha sido ampliamente debatido tanto en la tradición analítica como en la continental[53].

En segundo lugar, aun asumiendo una realidad mental que está generada por una determinada base biológica, no se muestra por qué el comportamiento de dicha base no es computable en su totalidad. Decir que no lo es porque los fenómenos mentales y el pensamiento tienen unas características –como la conciencia, la intencionalidad, la subjetividad, el sentido o el significado– que no se pueden generar por

48 Searle (1994), p. 27.
49 Žižek (2023), p. 94.
50 Larson (2022), p. 290-300.
51 Searle (1994), p. 40.
52 Puede verse al respecto las películas *Ex Machina*, dirigida por Alex Garland en 2014, y la más reciente *The Artifice Girl*, dirigida por Franklin Ritch en 2022.
53 No es objeto de este trabajo entrar en el debate. Al respecto cabe destacar, del lado de la tradición analítica, los siguientes estudios: Dennett (1995; 2017), Metzinger (2003), Damasio (2010). De la tradición continental destacamos: Rodríguez (1997), Xolocotzi (2002), Zahavi (2003).

procedimientos computacionales, es dar por supuesto lo que se quiere probar. Lo que Turing defiende en su artículo de 1950, cuando aborda este asunto discutiendo con el neurólogo Geoffrey Jefferson, es que 1) *a priori*, es decir, con independencia de la experiencia, no tenemos argumentos para afirmar que el comportamiento del cerebro y del sistema nervioso central no es computable; y 2) puesto que la relectura del teorema de Gödel deja abierta la posibilidad de generar sistemas computacionales cada vez más potentes fundamentados en los anteriores, hemos de atender al desarrollo de la computación para reconocer sus límites, que se harán patentes cuando encontremos un comportamiento cerebral, neuronal y nervioso que no sea computable. Insisto, en fin, en que Turing no defiende que la subjetividad y el pensamiento sean computables; lo que defiende es que no podemos negar que lo sean mientras no encontremos comportamientos cerebrales y nerviosos que no sean computables. Por ello repetía Turing que la respuesta a la pregunta sobre si las máquinas piensan es empírica, y que mientras no encontremos ese límite empírico nada impide defender que las máquinas que replican computacionalmente los comportamientos cerebrales y nerviosos son máquinas que piensan.

Quiero hacer hincapié en la sutileza del argumento del juego de la imitación cuando propone simular el comportamiento. Lo que se busca es una máquina que se comporte como un humano, de la misma manera que en 1942 se buscaba un computador que se comportase como Tunny. Bill Tutte ofrece a Turing un esquema del comportamiento de Tunny y fue éste el que en unas semanas ideó un procedimiento computable en papel para descifrar los mensajes de Tunny. El procedimiento es implementado por el ingeniero Tommy Flowers mediante válvulas electrónicas y procesos digitales, construyendo finalmente Coloso. Ninguno de ellos sabía cómo funcionaba Tunny, ni falta que hacía para tener una máquina computable que generaba efectos similares. De la misma manera, en 1950 no le interesa a Turing saber cómo funciona el cerebro, ni el sistema nervioso central, ni sus procesos físicos, eléctricos, químicos, moleculares, atómicos, etc. Todo ello exige una ardua investigación cuyo final no se atisba en el horizonte humano. Lo que le interesa es su comportamiento y, a partir de ahí, idear procedimientos computables que se puedan implementar en una máquina. Decir a priori que esos procedimientos computables no existen no está justificado. Se trata de ir probando y mejorando computacionalmente los procedimientos para ajustarlos al comportamiento humano. Como ya he dicho, la relectura del teorema de Gödel deja abierta la posibilidad de una mejora continua.

10. Sistemas computables que generan mente

Imaginemos un mundo en el que se ha alcanzado tal desarrollo de la computación que vivimos rodeados de máquinas que se comportan exactamente como humanos. El trato con unas y otros en nuestra vida cotidiana, que siempre acontece desde la perspectiva en tercera persona, sería idéntico y transcurriría sin preocuparnos de qué o quién tenemos delante. Sería irrelevante si funciona por medio de un sistema nervioso, eléctrico, químico, o completamente mecánico, como la máquina de Babbage. Lo importante, subraya Turing, es entender que las similitudes de comportamiento provienen de "analogías matemáticas de función"[54].

Pero no hablemos ahora de imitar el comportamiento cerebral o del sistema nervioso central, ni la acción humana, sino de generar una realidad mental. Un sistema computable que supera por completo el juego de la imitación y se comporta de manera idéntica a un humano, ¿sería capaz de desarrollar una realidad mental? ¿Podríamos afirmar la existencia de realidades mentales similares provenientes de analogías matemáticas de función?

Hemos visto que los autores referidos ofrecen una respuesta negativa, la cual se hace depender de una petición de principio. Turing se resiste a tratar el pensamiento desde los fenómenos mentales y nos remite en todo momento a la imitación de los comportamientos cerebrales y las acciones humanas. Por ello, comienza el artículo de 1950 sin dar una definición –mentalista– de pensamiento, y tampoco lo hace en las entrevistas que concedió posteriormente. Así deja abierta la posibilidad de que los sistemas computables generen fenómenos mentales y que, en definitiva, la subjetividad humana sea el producto de un cerebro y un sistema nervioso entendidos como sistemas algorítmicos recursivos. Dice Daniel Dennett que simplemente el hecho de dejar abierta tal posibilidad consigue "romper el hechizo" de Descartes[55], que fue el que nos hizo creer en dualismos. Turing no solo nos lleva a reconocer pensamiento en las máquinas de la misma manera que lo reconocemos en otros hombres, es decir, desde la perspectiva en tercera persona, sino que nos saca del dualismo ontológico cartesiano ofreciendo la posibilidad de entender las mentes como productos de sistemas computables equivalentes, es decir, sistemas que resultan ser analogías matemáticas de función. Fenómenos de inteligencia y comprensión surgirían a partir de meras competencias reguladas computacionalmente. "Turing –dice Dennett– abrió la puerta a un enorme espacio de diseño relativo al procesamiento de información y también previó que en ese espacio de diseño

54 Turing (1950/2004), p. 446.
55 Dennett (2017), cáp. 4.

había un camino transitable entre la ignorancia absoluta y la inteligencia artificial, un largo trecho de escalones ascendentes"[56]. Resulta posible que de la competencia sin comprensión de los algoritmos surjan los fenómenos de comprensión y sentido a los que asistimos en primera persona. Con esta idea regresa Turing a Cambridge en 1947, estudia los últimos desarrollos en biología y neurociencia y comienza a pensar en el crecimiento de formas biológicas a partir de ecuaciones no lineales, así como en el desarrollo de redes neuronales artificiales a partir de algoritmos recursivos[57]. Una parte de ese trabajo fue publicado en "La base química de la morfogénesis", del año 1952.

Así las cosas, respecto del argumento de la habitación china habría que decir lo que reconoce Arana, a saber, que "en contra de lo que asevera, Searle *acaba aprendiendo chino* dentro de la habitación"[58]. La persona que no sabe chino responderá a los mensajes que le llegan escritos en chino utilizando manuales, gramáticas y diccionarios, es decir, aplicando reglas, de la misma manera que lo hace una máquina. Les reconoceremos la competencia de responder chino y poco a poco, con el tiempo y la suficiente capacidad de almacenamiento, o bien, como exige Turing a la máquina, con la necesaria velocidad de procesamiento, se establecerán relaciones a partir de las semejanzas y las coincidencias detectadas, aparecerán conjeturas que podrán ser corregidas tras fracasos ulteriores, se almacenarán los resultados exitosos y se aplicarán tentativamente en nuevas conversaciones, y así hasta que la máquina, o la persona comportándose como una máquina, supere el test de Turing del idioma en cuestión, "sin necesidad de apoyarse en las instrucciones del programa, sino única y exclusivamente en sus propios fantasmas. Porque, al fin y al cabo, ¿quién nos garantiza que «de verdad» conocemos nuestro propio idioma?" O dicho de otra manera: ¿acaso no es eso y solo eso comprender nuestro idioma? La sintaxis arrastra consigo la semántica y las fronteras entre una y otra desaparecen. Con el programa adecuado, una gran capacidad de almacenamiento y una alta velocidad de procesamiento, la competencia de responder chino acaba generando la comprensión del chino. A lo cual hay que añadir, subraya M. Rodríguez, que la comprensión no sucede de una vez por todas, y que antes hay semi-comprensión, y antes semi-semi-comprensión, a partir de una cascada de competencias sin comprensión que va generando comprensión a cuartos, comprensión a medias, etc.[59]

11. Algoritmos recursivos

Turing no dice que el comportamiento del cerebro y del sistema nervioso es computable, sino que no tenemos argumentos para decir que no lo es, dejando abierta la posibilidad de que lo sea y ofreciendo de esta manera una explicación de los fenómenos mentales a partir de sistemas computables. En 1950 afirmaba que la evaluación de esta posibilidad requería un desarrollo de la tecnología de la computación que no se produciría antes de finales de siglo[60]. Pero el proyecto mismo de construcción de máquinas computacionales cada vez más potentes hasta alcanzar máquinas "supercríticas" [61] –así las denomina el propio Turing– que generen fenómenos mentales, es decir, el proyecto de construcción de una inteligencia artificial, de alguna manera ya se había puesto en marcha con la publicación de "On Computable Numbers" en 1936, las nuevas definiciones de recursividad de Kleene y la demostración de Church de la equivalencia entre la función recursiva general de Herbrand-Gödel-Kleene, la máquina universal de Turing y las funciones lambda de Church[62].

En 1893, en el trabajo titulado *Was sind und was sollen die Zahlen?*[63], Dedekind había utilizado las funciones recursivas para definir los números naturales. En 1922 Skolem utiliza similar procedimiento recursivo para reconstruir la lógica de los *Principia* de Whitehead y Russell, logrando eliminar los cuantificadores existenciales al sustituirlos por funciones y haciendo desaparecer consecuentemente el tan discutido problema de la existencia. En el artículo de 1931 titulado "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I", Gödel expone sus teoremas de incompletitud y ofrece la siguiente definición de las funciones recursivas[64]:

> Una función numérica *f* se llama *recursiva primitiva* si hay una secuencia finita de funciones *f1, f2, ..., fn* que acaba con *f* y que tiene la propiedad de que cada función *fk* de la secuencia está recursivamente definida a

56 Dennett (2017), p. 61.
57 Copeland (2013), cáp. XI.
58 Arana (2015), p. 56.
59 Rodríguez (2021), pp. 165-170.
60 Turing (1950/2004), p. 459.
61 Ibid.
62 Puede verse al respecto el magnífico estudio de W. Sieg, "Mechanical Procedures and Mathemathical Experience", en George (1994).
63 Véase Dedekind (2012).
64 En Gödel (1931/1981), p. 64.

> partir de dos de las funciones precedentes o resulta de alguna de las funciones precedentes por sustitución[65], o, finalmente, es una constante o la función del siguiente, x + 1.

A Gödel le llegó la crítica que poco tiempo después hizo Herbrand de esta definición y vuelve sobre el asunto en el curso que impartió en Princeton en la primavera de 1934, publicado a partir de los apuntes de Kleene y Rosser con el título *On undecidable propositions of formal mathematical systems.* Reconoce el "sentido limitado" que había dado a la recursión y define "cualquier función recursiva" del siguiente modo[66]:

> Si *f* es una función desconocida y *g1, ..., gn* son funciones conocidas, las *gi* y *f* se sustituyen una en otra de los modos más generales y ciertos pares de las expresiones resultantes se igualan entonces si el conjunto resultante de ecuaciones de funciones tiene una y solo una solución para *f*, entones *f* es una función recursiva.

Señala Yuk Hui que es esta nueva definición la que subraya las tres características de la recursión[67]: 1) lo que puede ser reducido a funciones recursivas es *computable*; 2) se parte de una función simple pero se llega a una función compleja que describe un fenómeno que se trata como *emergencia*, y 3) actúa como una *caja negra* que resulta válida en la medida que logra producir un determinado efecto conocido. El fracaso de la recursión sería el fracaso de la computación, esto es, que la función recursiva no alcance su meta, que el proceso (recursivo, computable) no se detenga y que entre en un bucle infinito que agote la memoria o incluso destruya físicamente la máquina.

Esta idea de recursividad es la que se mantiene en "La base química de la morfogénesis" al explicar algorítmicamente la embriología química –inspirada en el análisis de las formas biológicas de D´Arcy Thompson– y la generación de patrones y formas como manchas, vetas y motas en las alas de las mariposas, las conchas de los moluscos o la piel de los tigres y los leopardos[68]. De esta manera, afirma Langton, la computación sobre algoritmos recursivos pasa a ser una importante "herramienta de laboratorio para el estudio de la vida, sustituyendo la variedad de incubadoras, placas de cultivo, microscopios, geles electroforéticos, pipetas, centrifugadoras y otra parafernalia variada"[69]. Se abre así un amplio proyecto de investigación sobre vida artificial cuyo "objetivo final ... sería crear «vida» en algún otro medio, idealmente un medio *virtual* donde la esencia de la vida haya sido abstraída de los detalles de su implementación en cualquier hardware particular"[70].

Pero lo que quiero destacar es que el estudio computacional de la vida que hace Turing, tratando de explicar la organización y los patrones de los seres vivos por medio de embriología química y la explicación de los números de Fibonacci, persigue desde el comienzo entender el cerebro y la emergencia del pensamiento. Así lo dice en la carta que dirige a J.Z. Young[71]:

> Realmente estoy haciendo esto ahora porque está cediendo más fácilmente al tratamiento. Creo que no está del todo desconectado del otro problema. La estructura del cerebro tiene que ser una que pueda lograrse mediante el mecanismo genético embriológico, y espero que esta teoría en la que estoy trabajando ahora pueda aclarar qué restricciones esto realmente implica. Lo que usted me cuenta sobre el crecimiento de las neuronas bajo estimulación es muy interesante al respecto. Sugiere medios por los cuales se podría hacer que las neuronas crezcan para formar un circuito particular, en lugar de llegar a un lugar particular.

Surge de esta manera con el propio Turing una nueva disciplina, la *cibernética*, capaz de integrar máquina, vida y pensamiento a partir de la recursión, que es el mecanismo que explica las novedades y las emergencias como productos de una contingencia domesticada[72]. Una cibernética que tiene como objetivo último reducir todos los comportamientos cerebrales a algoritmos recursivos en máquinas computacionales cada vez más potentes, más veloces y con más capacidad de almacenamiento, y explicar de esta manera la emergencia del pensamiento. En ese proyecto seguimos y seguiremos, porque mientras no encontremos un comportamiento que no sea computable, la posibilidad de generar una inteligencia artificial permanece. Y los que la niegan *a priori*, es decir, sin esperar al devenir de los tiempos, lo hacen dogmáticamente, apoyados en una petición de principio. Esto es, en definitiva, lo que Turing nos quiso decir en 1950.

---

65 "Más precisamente: por introducción de algunas de las funciones precedentes en los lugares argumentales de una de las funciones precedentes, por ejemplo, *fk (x1, x2) = fp (fq (x1, x2), fr (x2)), (p, q, r < k)*. No es necesario que todas las variables del lado izquierdo aparezcan también en el derecho".
66 En Gödel (1934/1981), p. 178.
67 Hui (2022), p. 164.
68 Copeland (2004), pp. 508-509.
69 Langton (1989), p. 1. La traducción es mía.
70 Texto de Langton referido en Copeland (2004), p. 507. La traducción es mía.
71 En Copeland (2004), p. 517. La traducción es mía.
72 Hui (2022), pp. 201-210.

12. Conclusión

En "Computing Machinary and Intelligence" Turing define el pensamiento desde el comportamiento, abandonando la perspectiva en primera persona. La razón es que tal forma de proceder es la utilizada para reconocer pensamiento en alguien distinto de uno mismo. Lo que Turing defiende es que este mismo criterio se aplique también para decidir si una máquina piensa o no.

En contra de Turing, Searle defiende que el pensamiento no se puede definir exclusivamente desde el comportamiento, o desde las competencias, porque el pensamiento conlleva, además, sentido y significado, o dicho de otra manera, competencia con comprensión. Puesto que el comportamiento se puede reducir a procedimientos computacionales –meramente formales o sintácticos– pero no el sentido y el significado –esto es, la semántica–, hemos de concluir que las máquinas no pueden pensar.

Pero preguntémonos: ¿qué son el sentido y el significado, qué es la semántica? Estos términos nos remiten, a juicio de Serle, a una realidad mental, a unos estados subjetivos, "tan reales y tan irreductibles como cualquier cosa del universo". De tal manera que aunque las máquinas de Turing y los humanos realicen los mismos comportamientos, en las máquinas están producidos por procedimientos exclusivamente formales y en los humanos por el pensamiento, constituido por una sucesión de estados subjetivos intencionales y conscientes. Las mismas competencias se dan, en el primer caso, sin comprensión, y en el segundo con comprensión. Esta referencia a una realidad mental que, además, se constituye en una unidad y es condición de posibilidad de la emergencia del mundo con sentido, así como de sus modificaciones, se mantiene en los argumentos que, contra la posibilidad de una inteligencia artificial, han esgrimido recientemente Žižek y Larson.

El problema es que el reconocimiento de esa realidad mental que constituye el sentido y el significado, la semántica y la comprensión, solo se puede realizar desde la perspectiva en primera persona. Así las cosas, el núcleo del argumento de Turing en "Computing Machinery" no resulta rebatido, porque su exigencia es la de determinar que una máquina piensa con los mismos criterios que determinamos que una persona piensa, a saber, atendiendo a la perspectiva en tercera persona, es decir, al comportamiento.

Pero el asunto va más allá: si la realidad mental solo se puede reconocer desde la perspectiva en primera persona, ¿por qué autores como Searle, Žižek o Larson afirman que existe en otros que no soy yo y que no existe en las máquinas de Turing? Respondo: porque todos ellos suponen I) que la realidad mental es un producto de la biología de nuestro cerebro y de nuestro sistema nervioso central, y Γ) que esa base biológica no es computable. Lo primero les lleva a reconocer estados mentales y pensamiento en todo lo que tenga una biología humana –reconocible desde la perspectiva en tercera persona–, y lo segundo les lleva a negar la posibilidad de pensamiento en las máquinas, que al fin y al cabo no son más que cacharros con sistemas computables.

La conclusión a la que llego en este trabajo es que Turing también mantiene I), pero niega Γ), e intento aclarar por qué lo hace. Pensemos en lo siguiente: ¿qué lleva a afirmar a estos autores que la biología humana no es computable? El único argumento que dan es que no puede serlo porque la realidad mental, que es producto de esa base biológica, contiene unas características (ya sabemos: sentido, significado, conciencia, intencionalidad, comprensión, etc.) que no son computables. Pero con ello se está cometiendo una petición de principio, se está dando por supuesto lo que se quiere probar, a saber, que la realidad mental no es computable.

A partir de aquí aclaro la sutil posición de Turing en las páginas finales de "Computing Machinery". No afirma que la base biológica humana es computable; lo que afirma es que no podemos decir que no lo es mientas no encontremos un comportamiento biológico que no se pueda reducir a un procedimiento computable. Por consiguiente, no podemos decir *a priori* que el pensamiento no es computable y las máquinas no piensan ni pensarán nunca. Por el contrario, se trata de iniciar una larga investigación para ir reduciendo comportamientos biológicos, cerebrales y nerviosos a procedimientos computables. Mientras no encontremos un comportamiento biológico no computable, sigue abierta la posibilidad de crear una base computacional con el mismo comportamiento que nuestra base biológica y que genere sus mismos productos, entre ellos el pensamiento. Surge así una disciplina, la cibernética, que el propio Turing inaugura mediante el desarrollo de algoritmos recursivos, cuyo objetivo es el de generar inteligencia artificial. Que se consiga o no es algo que tendremos que comprobar empíricamente.

13. Referencias bibliográficas

Arana, J. (2015): *La conciencia inexplicada. Ensayo sobre los límites de la comprensión naturalista de la mente*, Madrid, Biblioteca Nueva.

Cole, D. (2020): "The Chinese Room Argument", en *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/chinese-room/#3.

Damasio, A. (2010): *Self Comes to Mind. Constructing the Conscious Brain*, New York, Pamtheon Books.

Copeland, B.J. (ed.) (2004): *The Essential Turing*, Oxford, Clarendon Press.

Copeland, B.J. (2013): *Alan Turing. El pionero de la era de la información*, Madrid, Turner.
Dedekind, R. (2012): *Was Sind und was Sollen die Zahlen?*, New York, Cambridge University Press.
Dennett, D. (1995): *La conciencia explicada*, Barcelona, Paidós.
Dennett, D. (2017): *De las bacterias a Bach. La evolución de la mente*, Barcelona, Pasado y Presente.
George, A. (ed.) (1994): *Mathematics and Mind*, New York, Oxford University Press.
Gödel, K. (1931): "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme (*Monatshefte für Mathematik y Physik*, nº. 38, pp. 173-198), en K. Gödel, *Obras ompletes*, Madrid, Alianza Editorial, 1981, pp. 55-90.
Gödel, K. (1934): *On undecidable propositions of formal mathematical systems* (Institute for Advance Study, Princeton, New Jersey), en K. Gödel, *Obras ompletes*, Madrid, Alianza Editorial, 1981, pp. 151-182.
Hui, Y. (2022): *Recursividad y contingencia*, Buenos Aires, Caja Negra.
Langton, C.G. (1986): "Studying Artificial life with Cellular Automata", *Physica D: Nonlinear Phenomena*, vol. 22, issues 1-3, pp. 120-149.
Langton, C.G. (ed.) (1989): *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*, Redwood City, Addison-Wesley.
Larson, E.J. (2022): *El mito de la inteligencia artificial*, Shackleton Books.
Metzinger, T. (2003): *Being No One*, Cambridge, MIT Press.
Peirce, Ch. S. (1965): *Collected Papers*, Cambridge, Massachusetts, Harvard University Press.
Rivadulla, A. (2010): "Estrategias del descubrimiento científico. Abducción y preducción", en *Filosofia e História da Ciência no Cone Sul. Seleção de Trabalhos do 6º Encontro*, Campinas, Associação de Filosofia e História da Ciência do Cone Sul (AFHIC), pp. 120- 129.
Rivadulla, A. (2015): *Meta, método y mito*, Madrid, Trotta.
Rodríguez, R. (1997): *La transformación hermenéutica de la fenomenología. Una interpretación de la obra temprana de Heidegger,* Madrid, Tecnos.
Rodríguez González, M. (2021): *Filosofía de la mente*, Madrid, Ediciones Complutense.
Searle, J.R. (1992): *Intencionalidad. Un ensayo en la filosofía de la mente*, Madrid, Tecnos.
Searle, J. R. (1994): *Mentes, cerebros y ciencia*, Madrid, Cátedra.
Turing, A. (1937): "On Computable Numbers with an application to *Entscheidungsproblem*", *Proceedings of the London Mathematical Society*, s. 2, vol. 42, pp. 230-265.
Turing, A. (1938): *Systems of Logic based on Ordinals*, Princeton, Seeley G. Mudd Manuscript Library.
Turing, A. (1948): "Intelligent Machinery", en B.J. Copeland (ed.), *The Essential Turing*, Oxford, Clarendon Press, 2004, pp. 410-440.
Turing, A. (1950): "Computing Machinery and Intelligence" (*Mind*, 49, pp. 433-460), en B.J. Copeland (ed.), *The Essential Turing*, Oxford, Clarendon Press, 2004, pp. 442-464.
Turing, A. (1952): "The Chemical Basis of Morphogenesis", *Philosophical Transactions of the Royal Society of London*, Series B, Biological Sciences, B 237 (641), p. 37-72, https://doi.org/10.1098%2Frstb.1952.0012.
Valor Yébenes, J.A. (2024): "Alan Turing y el origen de la inteligencia artificial: la superación de la intuición", en *Naturaleza y Libertad*, nº. 18, pp. 13-57.
Xolocotzi, A. (2002): *Der Umgang als Zugang. Der hermeneutisch-phänomenologische Zugang zum faktischen Leben in den frühen Freiburger Vorlesungen,* Berlin, Duncker&Humblot.
Zahavi, D. (2003): *Husserl's Phenomenology*, Stanford, Stanford University Press.
Žižek, S. (2023): *Hegel y el cerebro conectado*, Barcelona, Paidós.