

LIMITES PARA LA DEFINICION DE UNA INTELIGENCIA ARTIFICIAL

Una vez finalizada esta panorámica de los sistemas cognitivos, cabe preguntarse: hasta dónde se puede avanzar por esta línea y dónde están los límites de las soluciones que aporta la inteligencia artificial. Como ya se habrá podido intuir, desde la presentación e introducción de este libro, y como se intentará demostrar a lo largo de este capítulo, los límites a una inteligencia artificial, planteados dentro del ámbito del proceso de la información, son necesariamente coincidentes con los límites de la ingeniería del software.

Debe hacerse notar, que desde el punto de vista de la ingeniería a partir del cual está escrito este libro, el objetivo fundamental de la inteligencia artificial es, como se ha dicho, construir sistemas que den servicio al hombre y le ayuden a solucionar sus problemas, utilizando para ello, sus mismos esquemas de razonamiento. Para alcanzar dicho objetivo, la inteligencia artificial puede utilizar también los esquemas de otras inteligencias diferentes a la humana, por ejemplo, la de la Naturaleza, en la línea de lo que plantean los algoritmos genéticos.

El intento de emular la mente humana o superarla, sólo por el hecho de hacerlo, puede dar lugar a entretenidas discusiones que, aparte de algunos frutos intelectuales colaterales, distraen esfuerzos de la labor investigadora y desarrolladora, a menudo más monótona, aunque más fructífera.

En cualquier caso, la respuesta a la pregunta de esta discusión: ¿puede una inteligencia artificial alcanzar o incluso superar a la mente humana?, debe plantearse en los términos de cuales son los límites a una inteligencia artificial, para posteriormente plantear si la inteligencia

humana está afectada por los mismos límites. Como el lector podrá ya imaginar, esta última pregunta quedará abierta en algún punto.

A raíz de esta discusión se tienen por tanto, tres cuestiones que hay que resolver:

- a) Primero demostrar la equivalencia, desde el punto de vista teórico, que no práctico, entre los límites de la inteligencia artificial y los de cualquier otro planteamiento del proceso automático de la información.
- b) Segundo, establecer los límites generales para el procesamiento automático de la información.
- c) Y, finalmente, discutir en qué medida los límites anteriores afectan al razonamiento humano.

13.1. Equivalencia de las máquinas de proceso de la información.

La informática estudia el proceso de transformación automática de la información. Los computadores son las máquinas capaces de realizar dicha transformación y están constituidos:

- a) necesariamente, por elementos físicos (mecánicos, eléctricos, electrónicos, magnéticos, ópticos, etc.) a los que en conjunto se les da el nombre de **hardware** (soporte físico).
- b) usualmente, por elementos lógicos (programas y datos) a los que en conjunto se les da el nombre de **software** (soporte lógico).

Desde un punto de vista estrictamente teórico, esto es, asumiendo que se dispone de los recursos de tiempo y memoria suficientes, cualquier computador puede emular a otro sin más que recubrirle de la capa de software

adecuado. En la práctica, aunque pueda resultar impensable, se puede emular un supercomputador sobre una máquina personal, siendo necesario abundantes recursos de memoria y sabiendo de antemano que los tiempos de respuesta serán muy prolongados. Dicha emulación podría hacerse a la inversa. Es por todos conocido que, comercialmente, se distribuye software de emulación para rentabilizar desarrollos realizados sobre otras máquinas, de hecho, puede darse la paradoja de máquinas ya obsoletas y desaparecidas sigan existiendo, virtualmente, emuladas con otras.

Por otra parte, toda la lógica que se implementa en software puede hacerse en hardware, y así se hacía en los primeros tiempos de la informática, por ejemplo, el ENIAC (Electronic Numerical Integrator and Computer) de la Universidad de Pennsylvania, se programaba mediante conexiones eléctrica e interruptores. De hecho, sólo razones económicas (flexibilidad, transportabilidad, reutilización, etc.) justifican la existencia del software y por tanto, la frontera entre el hardware y software puede establecerse allí donde el diseñador crea conveniente.

Todas estas afirmaciones, intuitivas a partir de un conocimiento mínimo sobre los computadores, fueron establecidos por Alan Turing en los años 30. Turing planteó una máquina abstracta (ideal) capaz de dar solución a todos aquellos problemas que se pueden resolver con lápiz y papel en un tiempo finito, esto es, todos aquellos problemas para los que existe un algoritmo que conduce a su solución, entendiendo por algoritmo una secuencia de operaciones que proporcionan la solución a un problema es un número finito de pasos.

Estas máquinas están formadas por un hardware ideal que puede leer y escribir sobre una cinta en la que la información se guarda en forma de unos y ceros o en cualquier otro tipo de codificación. La programación de una **máquina de Turing** consiste en un conjunto de reglas de la forma: $e_i L \rightarrow e_j E M$

que indica que: si la máquina está en el estado e_i y en la cinta se lee la información L , la máquina transitará al estado e_j escribiendo en la cinta la información E y realizando un movimiento M de su cabezal sobre la cinta. Por tanto, la máquina de Turing tiene un conjunto finito de estados ($e_1, e_2, \dots, e_i, \dots, e_f, \dots, e_m$) y de movimientos de la cinta bajo el cabezal de lectura/escritura (derecha, izquierda y parada).

Turing demostró que puede definirse una máquina U , a la que se conoce **máquina universal de Turing**, en cuya cinta de entrada recibe:

- en primer lugar información para emular a otra

máquina de Turing y a continuación,

- los datos que se habrían de pasar a la máquina original.

Esta máquina universal U será capaz de presentar a su salida la misma solución que la máquina original. Con lo que volvemos a la idea intuitiva inicial de que toda máquina es equivalente a otra si se la recubre con el software necesario y que todo software sea cual sea su método de construcción puede introducirse como programa de una máquina de Turing, pues es un algoritmo.

Cualquier aplicación de inteligencia artificial, sean cuales sean sus técnicas de construcción y herramientas en las que se apoye, se ejecutará finalmente sobre un computador (todo computador general es una máquina universal de Turing) como una secuencia de instrucciones de ensamblador o microcódigo y por tanto, siempre cabría pensar que un buen programador de ensamblador podría haberlas construido directamente en código máquina y, por tanto, las limitaciones de la inteligencia artificial están allí donde estén las limitaciones del software. Como ya se ha comentado, la utilización de unas técnicas u otras frente a ciertas clases de problemas, deben elegirse por razones de facilidad de construcción, legibilidad, flexibilidad de modificación, capacidades de explicación, etc., esto es, sólo por razones económicas.

13.2. Límites del software

Pueden enunciarse tres limitaciones al proceso de la información. Como en el apartado anterior una demostración y explicación rigurosa de ellos se sale fuera de los objetivos de este libro.

Límite de lo universal

Al primero se le denominará **límite de lo universal**, en razón a su relación con la cantidad de información del universo. Este límite viene impuesto por la teoría de la información y es conocido como **tesis de Wiener**. La tesis de Wiener afirma que ninguna máquina de proceso de información puede presentar a su salida más información de la que recibe:



Figura 13.2.1: Proceso de información.

Esto es, para toda máquina M de proceso de información, si se le introduce una cantidad de información E (por ejemplo, en forma de programas y

datos), la cantidad de información S que dicha máquina presenta a su salida cumple que $S \leq E$. En otras palabras, ninguna máquina de proceso de la información puede aumentar la cantidad de información del universo. Intuitivamente, puede verse que a partir de unos datos de entrada, por ejemplo, 3 y 3, y de una descripción de cómo hacer una suma, programa, una máquina puede presentar a su salida un 6, pero la información que aporta este 6 está incluida en los dos treses y en el concepto de suma, otra cosa es que algún observador externo desconociera que $3 + 3 = 6$. De la misma forma, un sistema experto en clasificación de piezas arqueológicas, por ejemplo, puede, aparentemente, aportar nuevos datos sobre una moneda romana, pero dicha información ya está incluida en las características de la moneda y en los criterios heurísticos de los arqueólogos que aportaron su conocimiento durante la construcción del sistema experto.

Límite de lo humano

Al segundo límite se lo denominará **límite de lo humano**, con un doble sentido que corresponde tanto a si en él está el límite al que puede llegar el razonamiento humano o en el está el límite máximo en el que el computador puede parecerse al humano. Este límite viene impuesto por la teoría de la computabilidad y es conocido como **tesis de Church-Turing**. Se ha visto que cualquier aplicación de proceso de la información puede ser realizado por una máquina de Turing, es más, resulta matemáticamente razonable definir que una operación algorítmica es aquella que puede ser realizada por una máquina de Turing, los problemas que pueden ser resueltos por una máquina de Turing se denominan computables, pero existen muchos problemas que no lo son, problemas no computables.

Turing, originalmente, desarrolló sus ideas para dar solución a un problema planteado por David Hilbert a principios de siglo: ¿existe algún procedimiento automático para resolver todos los problemas matemáticos pertenecientes a una clase amplia bien definida?. Turing planteó este problema en términos de su máquina. Dada cierta máquina M y un problema de entrada P , dará la máquina una solución, esto es, parará tras un cierto tiempo de proceso (dando lugar a que se conociera al problema como él de parada de una máquina de Turing). Turing demostró que no existe forma de responder en general a dicha cuestión, que no existe un algoritmo universal para decidir si una máquina de Turing se parará o no. Y por tanto, existe toda una clase de problemas, cuya solución no es algorítmica, para los que no se puede construir un programa que les de solución sobre un computador, sean cuales sean las técnicas que se utilicen, entre ellas las de inteligencia

artificial.

Se considera, siendo generalmente aceptado, que la respuesta de Alan Turing a David Hilbert ya estaba recogida por Kurt Gödel en el teorema que lleva su nombre "sobre proposiciones formalmente indecidibles en los Principia Mathematica" [Gödel, 31], como respuesta a los Principia Mathematica de Bertrand Russell y Alfred North Whitehead. La Proposición IV de [Gödel, 31] podría enunciarse de forma simplificada como:

"toda formulación axiomática de teoría de los números incluye proposiciones indecidibles" traducción de [Hofstadter, 87].

Las máquinas de Turing, y por tanto los computadores, son una formulación axiomática de teoría de los números, y por tanto, incluyen proposiciones indecidibles, como por ejemplo la solución general al problema de parada.

Límite de lo físico

Al último límite se le denominará **límite de lo físico**, este límite viene impuesto por la teoría de la relatividad y es conocido como límite cuántico. Evans, tomando como elemento de referencia el número de neuronas de un cerebro humano (diez mil millones) realizó la siguiente sobre tamaños y consumos:

- Utilizando tecnología de válvulas (1945) sería necesario el tamaño de la ciudad de Nueva York y el consumo en potencia de su metro.
- Con tecnologías de transistores (1960) el tamaño sería el de la estatua de la libertad con un consumo de diez mil vatios.
- Utilizando alta escala de integración (LSI, Large Scale Integration, 1970), el volumen sería el de un autobús y el consumo de cientos de vatios.
- Con muy alta escala de integración (VLSI, Very Large Scale Integration, 1980), el tamaño sería el de un televisor con un consumo de potencia de unos pocos vatios.

Históricamente, esta carrera por reducir el consumo potencia aumentando, simultáneamente, la velocidad de proceso ha dado resultados espectaculares. Pero alcanzadas frecuencias de Gigaherzios el aumento de la velocidad de conmutación de los dispositivos se ha visto frenada. Las tecnologías basadas en Arseniuro de Galio, los superconductores, los dispositivos ópticos basados en el interferómetro de Fabry-Perot, aun superándose unos a otros, se enfrentan, todos ellos, con un límite cuántico infranqueable, no se puede aumentar la velocidad de conmutación a la vez que disminuir la potencia disipada de forma indefinida. Y de aquí la necesidad de investigar

en nuevas arquitecturas de computadores con paralelismo de datos y/o instrucciones. Se trata, por tanto, de un límite práctico, dado que la calidad de las máquinas, como función de sus alta velocidad y bajo consumo, esta limitada físicamente, habrá problemas que, aunque se conozcan los mejores y más rápidos métodos para su solución, no se podrá disponer jamás de la máquina de procese su solución en un tiempo razonable.

Estos tres límites se muestran en la figura 13.2.2, el área rallada representa todo aquellos problemas solucionables con un computador. En cualquier caso el conjunto de problemas cuya solución no esta limitada por ninguna de los tres es lo suficientemente amplio como para que un profesional no los note, muy a menudo, en su trabajo. De forma que en vez de desanimar deben servir para enmarcar correctamente los caminos que se han de seguir en la investigación y en el desarrollo.

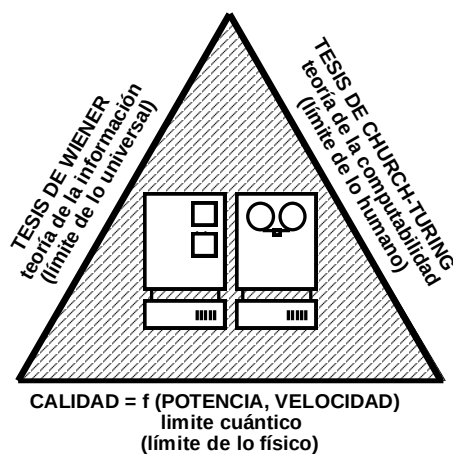


Figura 13.2.2: Límites del software.

13.3. Límites de la inteligencia natural.

Si las inteligencias naturales se vieran afectadas por los tres límites anteriores, entonces sería planteable que, alguna vez, se pudiera construir una inteligencia artificial que las igualara o superara en algunos casos. Esto abre dos cuestiones: la primera, ya planteada anteriormente, es que en qué medida está limitada la inteligencia natural y una nueva, cuando se crea haber construido una inteligencia artificial, cómo comprobar que es realmente una inteligencia.

Razonar en que medida los tres límites anteriores afectan a la inteligencia natural, afectan al hombre, entra dentro del campo de la filosofía. Tomar un partido radical por alguna de ellas no es el objetivo de este libro, ni mucho menos azuzar una discusión que se enmarque dentro del síndrome del doctor Frankenstein:

- En principio, el hombre como procesador de información, está limitado por la tesis de Wiener. Puede descubrir nuevos aspectos del universo que desconocía, pero no aportar nueva información que aumente la cantidad de esta en el universo. Aunque esta afirmación a veces resulte difícil de creer, por ejemplo ante la creación artística.
- El segundo límite, tesis de Church-Turing implicaría que la inteligencia humana es determinista, es equivalente a una máquina de Turing. Es la limitación que suele parecer, radicalmente, más falsa. Parece difícil que procesos tales como el amor y los sentimientos sean generados por procesos algorítmicos. Sin embargo, si se consideran los componentes físicos del razonamiento humano, todavía no del todo conocidos, como elementos de comunicación y elaboración de información, entonces, ¿por qué no creer que existe una máquina de Turing equivalente?. Turing solía afirmar que la inteligencia estaba afectada por un síndrome de masa crítica, esto es, a partir de un cierto número de componentes, muy elevado, el comportamiento de un computador empieza a ser impredecible, ni siquiera por sus programadores, y ello le dota de un aspecto inteligente.
- Finalmente, nuestros elementos básicos de razonamiento, las neuronas, están afectadas por el mismo límite cuántico, se distinguen de otros elementos artificiales por consumir muy poca potencia pero, a su vez, tener una baja velocidad de conmutación. Las características del cerebro, como conjunto de neuronas, parecen provenir, más que de la calidad de sus elementos básicos, de su compleja arquitectura y de sus inmensa capacidad para realizar procesamiento paralelo, tanto en cuanto a datos como instrucciones, capacidad que las arquitecturas conexionistas de computadores tratan de emular.

Las posturas frente a estos límites podrían organizarse en tres categorías, apareciendo, naturalmente, combinaciones de ellas:

- Existen razones espirituales que establecen una radical barrera entre el ser humano y las máquinas y esta barrera es, obviamente infranqueable. Ante este argumento no cabe plantearse ningún otro tipo de discusión, se cree o no se cree.
- Existen razones físicas, intuitivas y aun no descubiertas, por las cuales la inteligencia natural puede alcanzar a resolver problemas que las máquinas no pueden. Sostener este punto de vista

implica sostener que dichas potencialidades, que se basan en características y leyes físicas, no podrán ser jamás trasladables a una máquina.

- c) La inteligencia natural esta afectada por los tres límites anteriores y entonces es esperable construir una inteligencia artificial equivalente.

En cualquiera de los tres casos, y tras una discusión, basta con encender las máquinas y poner en marcha las aplicaciones, incluso las de sistemas cognitivos, en las que se han invertido tantos años-hombre, para comprobar, irremediablemente, sea cual sea la respuesta a las cuestiones anteriores, cuan lejos se está de disponer de inteligencias artificiales, que aun con reparos puedan ser calificadas, mínimamente, de inteligentes. Aunque, por otra parte, suele decirse que procesos para cuya solución parece necesaria la utilización de la inteligencia, en cuanto se consiguen mecanizar pierden, en apariencia, esta cualidad.

13.4. Test de inteligencia.

Fuera ya de la discusión teórica anterior, sean las inteligencias naturales o artificiales equivalentes o no, se podría marcar una línea empírica. Si se dispone de un test para discernir cuándo un ente es inteligente de cuándo no lo es, se podría asumir que todo sistema artificial es inteligente cuando supere dicho test. [Turing, 50] propuso un test que lleva su nombre (test de Turing) para realizar dicha prueba. Este test se podría enunciar de la forma siguiente: frente a un observador externo se sitúa un ser humano y una máquina que pretende ser inteligente, los aspectos externos, no indicativos de inteligencia, deberán de ocultarse (aspecto físico, tono de voz, olor corporal, etc.) se podrá dar el título de inteligente a la máquina si, tras mantener una comunicación con ambos, el observador externo no tiene criterios para saber cual de ellos posee una inteligencia natural. [Turing, 50] predijo que cerca del año 2000 habría máquinas que superarían el test.

Debe observarse que más que un test de inteligencia es un test de imitación, pues la máquina debe mostrar un razonamiento parecido al humano no sólo en sus capacidades sino, también, en sus limitaciones. En el momento que la máquina respondiera rápida y exactamente a una multiplicación de dos números de siete cifras sería descubierta. Por otra parte cabe pensar, que cierto programa pueda engañar a más gente que otro programa, no implica, necesariamente, que el segundo sea menos inteligente. Ya durante los años sesenta se desarrollaron programas cuyo objetivo era única y exclusivamente dar una apariencia engañosamente humana, son muy conocidos ELIZA de J.

Weizenbaum o PARRY de Colby.

Desde 1991, hay anualmente un concurso mundial, el premio Loebner, basado en este test. A través de ocho terminales seis máquinas y dos humanos se comunican con diez jueces que deben puntuar a cada terminal según su calidad de conversación humana. Los jueces sólo saben que tras los terminales hay, al menos, dos humanos y dos máquinas. Gana el premio anual la máquina que consiga engañar a más jueces. En 1991, la conversación se mantenía sobre un tema concreto y una de las máquinas consiguió engañar a cinco de los diez jueces, y sorprendentemente tres de los jueces creyeron que uno de los humanos era una máquina [Johnson, 92]. Cuando una máquina consiga ganar se le concederá el gran premio, que está todavía desierto, y el concurso se cerrará.

Anticipándose a la aparición de máquinas que superen ampliamente el test de Turing, [Searl, 80] propuso un argumento muy comentado, tanto dentro del mundo de la inteligencia artificial como fuera de él, este argumento es conocido como la oficina china, una oficina donde se ofrecía un excelente servicio de traducción en chino, pero cuyos trabajadores utilizaban un manual perfecto de traducción que les indicaba como cambiar unos símbolos por otros y desconocían en todo momento el significado de lo que traducían. La argumentación se basa en que aunque una máquina pueda aparentar un razonamiento inteligente funciona como un simple manipulador de símbolos desconociendo, absolutamente, sus significados. Mientras que el humano conoce los significados de los símbolos que manipula. Sin embargo, se podría argumentar que cuando los humanos dicen conocer los significados los expresan a través de otros símbolos de los que también creen conocer su significado. Símbolos que remiten a símbolos en forma de inmensas redes referenciales al final de las cuales siempre se parecen vislumbrar significados, pero que cualquier intento de navegación hacia el final de la red muestra un nuevo mar de símbolos que remiten a símbolos, ¿puede ser aplicado el argumento de Searle en contra del propio razonamiento humano?, este contra-argumento es parecido a la trampa que la Tortuga tiende a Aquiles en un diálogo de Lewis Carroll.

Ser conocedor del significado de lo que se manipula remite a una pregunta más trascendente, son las máquinas por inteligentes que aparezcan conscientes de si mismas, puede parecer que no, pero en cualquier caso, si a alguien le negamos que tenga conciencia de si mismo, ¿puede ese alguien hacer algo para convencer de que si la tiene?, sólo puede afirmarlo y los demás creerle o no. Turing ya se adelantó a este problema, una persona sólo puede estar segura de su propia

consciencia y a los demás concederles o no el beneficio de la duda, y si se concede dicho beneficio ¿por qué no dárselo también a una máquina que escriba en su terminal: "soy consciente de mi misma"?

El problema que subyace detrás de toda esta discusión es que no hay buenas pruebas de inteligencia, es más que ni siquiera hay una clara definición de lo que es la inteligencia, hasta mediados de este siglo sólo los humanos poseían esa característica por lo cual siempre se podía recurrir a una definición restrictiva, inteligencia es esa cualidad que poseen los humanos y que no poseen el resto de seres y cosas que les rodean, de hecho el test de Turing se basa en ello, pero ahora que pueden aparecer máquinas que aspiran a dicho adjetivo, se hace necesaria su definición.

[Newell, 90] da una definición de inteligencia, que no siendo, a lo mejor, válida para el mundo en general, puede ser muy útil para los ingenieros constructores de sistemas cognitivos:

"...un sistema es inteligente en la misma medida que se aproxima al nivel del conocimiento, y por tanto, inteligencia es la capacidad de poner los conocimientos al servicio de los objetivos que el sistema debe alcanzar..."

De esta forma un sistema cognitivo definido al nivel del conocimiento será inteligente si opera bajo el principio de racionalidad y utiliza, en cada momento, aquellos conocimientos que resulten más adecuados para conseguir los objetivos que se pretenden con dicho sistema. Por tanto, no importa que un sistema cognitivo para resolver ciertas clases de problemas no se muestre inteligente para resolver problemas que están fuera de sus objetivos de diseño o que ni siquiera los pueda resolver. Usualmente, se entiende que la inteligencia es para algo, y de esta forma la definición de Newell es aplicable en la realidad, se puede comprobar, sin que se considere una contradicción, que ante ciertas tareas algunos humanos parecen más inteligentes que otros y ante otras circunstancias esas apariencias se intercambian. Por ejemplo, en castellano se utiliza el adjetivo listo para aquellas personas que muestran gran inteligencia en lo cotidiano, y también se dice que cierta persona es muy inteligente, por que es capaz de resolver problemas muy abstractos, y a continuación se puntualiza que no es muy listo porque siempre acaba pagando el café.

Desde esta perspectiva y en el extremo, la definición de Newell permite afirmar que un sistema software es inteligente si está bien definido (nótese que la definición al nivel del conocimiento de Newell tiene una gran analogía con el diseño clásico de software). Un pequeño

programa, cuyos objetivos sean sumar y restar números, es inteligente si para sumar utiliza la operación de suma y para restar la de resta.

Bibliografía recomendada.

[Hofstadter, 87] Hofstadter, D.R.: "Gödel, Escher, Bach: Un eterno y grácil bucle" en Tusquets Editores & Consejo Nacional de Ciencias y Tecnología de México, 1987.

[Searl, 80] Searl, J.R.: "Minds, brains and programs", Behavioral and Brain Sciences, Vol.3, pages 417-424, 1980.

[Turing, 63] Turing, A.M.: "Computing machinery and intelligence" en Computers and Thought (a collection of articles) edición al cuidado de Feigenbaum, E.A., Feldman, J. (reedición de [Turing, 50]), Mc Graw-Hill Book Company, 1963.