

INTELIGENCIA ARTIFICIAL

A lo largo de la historia se han seguido cuatro enfoques para definir la Inteligencia artificial. Como es de esperar, existe un enfrentamiento entre los diferentes enfoques. Tenemos dos enfoques centrados en los humanos, y otros dos centrados en torno a la racionalidad. El enfoque centrado en el comportamiento humano, mide el éxito en términos de la fidelidad en la forma de actuar de los humanos, mientras que los enfoques centrados en la racionalidad toman como referencia un concepto ideal de la inteligencia, que llamaremos racionalidad. Un sistema racional hace lo “correcto”, en función de su conocimiento. A continuación describimos diferentes definiciones de la Inteligencia artificial que aparecen en libros.

Sistemas que piensan como humanos	Sistemas que piensan racionalmente
<p>“El nuevo y excitante esfuerzo de hacer que los computadores piensen ... máquinas con mentes, en el más amplio sentido literal” (Haugeland, 1985)</p> <p>“La automatización de actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas, aprendizaje ...” (Bellman, 1978)</p>	<p>“El estudio de las facultades mentales mediante el uso de modelos computacionales” (Charniak y McDermott, 1985)</p> <p>“El estudio de los cálculos que hacen posible percibir, razonar y actuar” (Winston, 1992)</p>
Sistemas que actúan como humanos	Sistemas que actúan racionalmente
<p>“El arte de desarrollar máquinas con capacidad para realizar funciones que cuando son realizadas por personas requieren inteligencia” Kurzweil, 1990)</p> <p>El estudio de cómo lograr que los computadores realicen tareas que, por el momento, los humanos hacen mejor” (Rich y Knight, 1991)</p>	<p>La inteligencia computacional es el estudio del diseño de agentes inteligentes” (Poole et al., 1998)</p> <p>“IA... está relacionada con conductas inteligentes en artefactos” (Nilsson, 1998)</p>

Conviene aclarar que al distinguir entre comportamiento humano y racional no se está sugiriendo que los humanos son necesariamente irracionales en el sentido de la inestabilidad emocional o desequilibrio mental. Basta con darnos cuenta de que no somos perfectos: no todos somos maestros de ajedrez, incluso aquellos que conocen todas las reglas del ajedrez; y desafortunadamente, no todos obtenemos un sobresaliente en un examen.

Comportamiento humano: el enfoque de la Prueba de Turing

La Prueba de Turing, propuesta por Alan Turing (1950), se diseñó para proporcionar una definición operacional y satisfactoria de inteligencia. En vez de proporcionar una lista larga y quizá controvertida de cualidades necesarias para obtener inteligencia artificialmente, Turing sugirió, una prueba basada en la incapacidad de diferenciar entre entidades inteligentes indiscutibles y seres humanos. El computador supera la prueba si un evaluador humano no es capaz de distinguir si las respuestas, a una serie de

preguntas planteadas, son de una persona o no. Hoy por hoy, podemos decir que programar un computador para que supere la prueba requiere un trabajo considerable. El computador debería poseer las siguientes capacidades:

- **Procesamiento de lenguaje natural** que le permita comunicarse satisfactoriamente en algún idioma.
- **Representación del conocimiento** para almacenar lo que se conoce o siente.
- **Razonamiento automático** para utilizar la información almacenada para responder a preguntas y extraer nuevas conclusiones.
- **Aprendizaje automático** para adaptarse a nuevas circunstancias y para detectar y extrapolar patrones.

La Prueba de Turing evitó deliberadamente la interacción física directa entre el evaluador y el computador, dado que para medir la inteligencia es innecesario simular físicamente a una persona. Sin embargo, la llamada Prueba Global de Turing incluye una señal de vídeo que permite al evaluador valorar la capacidad de percepción del evaluado, y también le da la oportunidad al evaluador de pasar objetos físicos a través de una ventanita. Para superar la Prueba Global de Turing el computador debe estar dotado de:

- **Visión computacional** para percibir objetos.
- **Robótica** para manipular y mover objetos.

Éstas seis disciplinas abarcan la mayor parte de la IA, y Turing merece ser reconocido por diseñar una prueba que se conserva vigente después de 50 años. Los investigadores del campo de la IA han dedicado poco esfuerzo a la evaluación de sus sistemas con la prueba de Turing, por creer que es más importante el estudio de los principios en los que se basa la inteligencia que duplicar un ejemplar. La búsqueda de un artefacto que “volara artificialmente” tuvo éxito cuando los hermanos Wright, entre otros, dejaron de imitar a los pájaros y comprendieron los principios de la aerodinámica. Los textos de ingeniería aerodinámica no definen el objetivo de su campo como la construcción de máquinas que vuelen como palomas de forja que puedan incluso confundir a otras palomas.

Pensar como un humano: el enfoque del modelo cognitivo

Para poder decir que un programa dado piensa como un humano, es necesario contar con un mecanismo para determinar cómo piensan los humanos. Es necesario penetrar en el funcionamiento de las mentes humanas. Hay dos formas de hacerlo: mediante introspección (intentando atrapar nuestros propios pensamientos conforme éstos van apareciendo) y mediante experimentos psicológicos. Una vez se cuente con una teoría lo suficientemente precisa sobre cómo trabaja la mente, se podrá expresar esa teoría en la forma de un programa de computador. Si los datos de entrada/salida del programa y los tiempos de reacción son similares a los de un humano, existe la evidencia de que algunos de los mecanismos del programa se pueden comparar con los que utilizan los seres humanos. Por ejemplo, a Allen Newell y Herbert Simon, que desarrollaron el “Sistema de Resolución General de Problemas” (SRGP) (Newell y Simon, 1961), no les bastó con que su programa resolviera correctamente los problemas propuestos. Lo que les interesaba era seguir la pista de las etapas del proceso de razonamiento y compararlas con las seguidas por humanos a los que se les enfrentó a los mismos problemas. En el campo interdisciplinario de la ciencia cognitiva convergen modelos computacionales

de IA y técnicas experimentales de psicología intentando elaborar teorías precisas y verificables sobre el funcionamiento de la mente humana.

En los comienzos de la IA había confusión entre las distintas aproximaciones: un autor podría argumentar que un algoritmo resolvía adecuadamente una tarea y que por tanto era un buen modelo de representación humana, o viceversa. Los autores actuales hacen diferencia entre las dos reivindicaciones; esta distinción ha permitido que ambas disciplinas, IA y ciencia cognitiva, se desarrollen más rápidamente. Los dos campos continúan alimentándose entre sí, especialmente en las áreas de la visión y el lenguaje natural.

Pensamiento racional: el enfoque de las leyes del pensamiento

El filósofo griego Aristóteles fue uno de los primeros en intentar codificar la “manera correcta de pensar”, es decir, un proceso de razonamiento irrefutable. Sus silogismos son esquemas de estructuras de argumentación mediante las que siempre se llega a conclusiones correctas (por ejemplo: “Sócrates es un hombre; todos los hombres son mortales; por lo tanto Sócrates es mortal”). Estas leyes de pensamiento supuestamente gobiernan la manera de operar de la mente; su estudio fue el inicio de un campo llamado lógica.

Estudiosos de la lógica desarrollaron, en el siglo XIX, una notación precisa para definir sentencias sobre todo tipo de elementos del mundo y especificar relaciones entre ellos (compárese esto con la notación aritmética común, que prácticamente sólo sirve para representar afirmaciones acerca de la igualdad y desigualdad entre números). Ya en 1965 existían programas que, en principio, resolvían cualquier problema resoluble descrito en notación lógica. La llamada tradición logista dentro del campo de la inteligencia artificial trata de construir sistemas inteligentes a partir de estos programas.

Este enfoque presenta dos obstáculos. No es fácil transformar conocimiento informal y expresarlo en los términos formales que requieren de notación lógica, particularmente cuando el conocimiento que se tiene es inferior al 100 por 100. En segundo lugar, hay una gran diferencia entre poder resolver un problema “en principio” y hacerlo en la práctica. Incluso problemas con apenas una docena de datos pueden agotar los recursos computacionales de cualquier computador a menos que cuente con alguna directiva sobre los pasos de razonamiento que hay que llevar a cabo primero. Aunque los dos obstáculos anteriores están presentes en todo intento de construir sistemas de razonamiento computacional, surgieron por primera vez en la tradición lógica.

Actuar de forma racional: el enfoque del agente racional

Los agentes informáticos se espera que tengan atributos que los distingan de los “programas” convencionales, como que estén dotados de controles autónomos, que perciban su entorno, que persistan durante un período de tiempo prolongado, que se adapten a los cambios, y que sean capaces de alcanzar objetivos diferentes. Un agente racional es aquel que actúa con la intención de alcanzar el mejor resultado o, cuando hay incertidumbre, el mejor resultado esperado.

En el caso del enfoque de la IA según las “leyes del pensamiento”, todo el énfasis se pone en hacer inferencias correctas. La obtención de estas inferencias correctas puede, a veces, formar parte de lo que se considera un agente racional, ya que una manera racional de actuar es llegar a la conclusión lógica de que si una acción dada permite alcanzar un objetivo, hay que llevar a cabo dicha acción. Sin embargo, el efectuar una inferencia correcta no depende siempre de la racionalidad, ya que existen situaciones para las que no hay nada correcto que hacer y en las que hay que tomar una decisión. Existen también formas de actuar racionalmente que no implican realizar inferencias. Por ejemplo, el

retirar la mano de una estufa caliente es un acto reflejo mucho más eficiente que una respuesta lenta llevada a cabo tras una deliberación cuidadosa.

Todas las habilidades que se necesitan en la Prueba de Turing deben permitir emprender acciones racionales. Por lo tanto, es necesario contar con la capacidad para representar el conocimiento y razonar basándonos en él, porque ello permitirá alcanzar decisiones correctas en una amplia gama de situaciones. Es necesario ser capaz de generar sentencias comprensibles en lenguaje natural, ya que el enunciado de tales oraciones permite a los agentes desenvolverse en una sociedad compleja. El aprendizaje no se lleva a cabo por erudición exclusivamente, sino que profundizar en el conocimiento de cómo funciona el mundo facilita la concepción de estrategias mejores para manejarse en él.

La percepción visual es necesaria para poder tener una idea mejor de lo que una acción puede llegar a representar, por ejemplo, el ver un delicioso bocadillo contribuirá a que nos acerquemos a él.

Por esta razón, el estudiar la IA desde el enfoque del diseño de un agente racional ofrece al menos dos ventajas. La primera es más general que el enfoque que proporcionan las “leyes del pensamiento”, dado que el efectuar inferencias correctas es sólo uno de los mecanismos existentes para garantizar la racionalidad. La segunda es más afín a la forma en la que se ha producido el avance científico que los enfoques basados en la conducta o pensamiento humano, porque la norma de la racionalidad está claramente definida y es de aplicación general. Por el contrario, la conducta humana se adapta bien a un entorno específico, y en parte, es producto de un proceso evolutivo complejo, en gran medida desconocido, que aún está lejos de llevarnos a la perfección.

Un elemento importante a tener en cuenta es el siguiente: más bien pronto que tarde se observará cómo obtener una racionalidad perfecta (hacer siempre lo correcto) no es posible en entornos complejos. La demanda computacional que esto implica es demasiado grande. En la mayor parte de esta obra se adoptará la hipótesis de trabajo de que la racionalidad perfecta es un buen punto de partida para el análisis. Lo cual simplifica el problema y proporciona el escenario base adecuado sobre el que se asientan los cimientos de este campo.