

## APÉNDICE

### INTELIGENCIA ARTIFICIAL

Esta novedosa área de investigación enfoca sus esfuerzos hacia la comprensión y construcción de agentes inteligentes, en otras palabras, a dotar a las computadoras de un nivel humano de inteligencia.

Se entiende por la expresión “agente” todo aquello que puede considerarse que percibe su ambiente mediante sensores y que responde o actúa en tal ambiente mediante efectores.

Los agente biológicos como los humanos, cuentan con ciertos órganos o sentidos que les permiten recabar información del medio y responder a sus exigencias. En el caso de los agentes de *software*, sus percepciones y acciones equivalen a cadenas de bits codificados.

No existe acuerdo generalizado acerca de las facultades cognitivas que quedarían denotadas por la expresión “inteligencia”, sin embargo podemos mencionar algunas que sin duda son consideradas como expresiones o manifestaciones de ella tales como aprender, recordar, razonar, ver, etcétera.

El surgimiento de las computadoras permitió la transición del estado especulativo en el que se encontraban los estudios sobre la inteligencia humana iniciados hace más de 2000 años, hacia la consolidación de una disciplina teórica y experimental que hace posible que se prueben empíricamente las diversas teorías relacionadas con facultades cognitivas.

Esta disciplina ha servido como punto de encuentro de científicos interesados en diversos aspectos del intelecto desde múltiples puntos de vista como la psicología, la lingüística, la filoso-

fía de la mente, etcétera. De ahí su carácter eminentemente interdisciplinario.

Esta situación permite por una parte que la inteligencia artificial o IA se beneficie de los resultados obtenidos en otros campos mediante, por ejemplo, la importación de sus metodologías de estudio, y por otra, el vocabulario y herramientas propias de la IA permite a las otras disciplinas sistematizar, integrar y automatizar sus conocimientos.

Russell apunta cuatro tendencias generales que guían la construcción de agentes inteligentes, las cuales son:

- a) Pensar como humano,
- b) Actuar como humano,
- c) Pensar racionalmente, y
- d) Actuar racionalmente.

Las tendencias *a)* y *c)* hacen énfasis en los procesos cognitivos internos, en el “pensamiento”, mientras que *b)* y *d)* enfatizan el comportamiento externo, es decir la manera en que actúa el agente.

Así también, *a)* y *b)* proponen como punto de comparación al humano, mientras que *c)* y *d)* a un ideal de racionalidad.

La construcción de agentes inteligentes que sigue la tendencia expuesta en *a)* se plantea como objetivo previo, desarrollar una teoría bastante precisa de la mente humana para su posterior expresión en un programa de computadora. Para efectos de la evaluación de dicho programa se toma en consideración la correspondencia de los datos de entrada/salida de la máquina, así como de la duración de su comportamiento, con la conducta “normal” de un humano.

Los agentes inteligentes cuya construcción sigue la tendencia expuesta en *b)*, tienen su origen en la prueba diseñada por Turing en 1950 basada en una definición operativa de la inteligencia.

Para Turing la inteligencia en una computadora se define como la capacidad de lograr eficiencia a nivel humano en todas las

actividades de tipo cognoscitivo, suficiente para engañar a un evaluador externo.

La prueba que diseñó (llamada “prueba de Turing”) consiste en contar con un evaluador (humano) quien realiza un interrogatorio en lenguaje natural a una máquina (sin que el evaluador sepa que se trata de una computadora). El éxito de la prueba se verifica si al término del interrogatorio, el evaluador no puede determinar si las respuestas han sido proporcionadas por una computadora o por un humano.

Para que una computadora pueda pasar la prueba de Turing, su programación requiere de lo siguiente:

- a) Que sea capaz de procesar un lenguaje natural (inglés, español, alemán, etcétera),
- b) Que sea capaz de representar el conocimiento,
- c) Que sea capaz de razonar automáticamente, y
- d) Autoaprendizaje de la máquina.

El enfoque expuesto en *c)* también llamado de las leyes del pensamiento, tiene su origen en Aristóteles quien realiza el primer intento de codificación de “la manera correcta de pensar”, es decir, de los procesos de pensamiento irrefutables a los que llamó “silogismos”, los cuales constituyen esquemas de estructuras argumentativas que garantizan el arribo a conclusiones verdaderas siempre que las premisas que las anteceden también lo sean.

Posteriormente con el desarrollo de la lógica formal durante los siglos XIX y XX (principalmente con los trabajos de Frege y Russell), se cuenta con una notación precisa para representar aseveraciones acerca de lo que existe en el mundo, así como sus relaciones. La tradición logicista en la IA se vale de estas aportaciones.

El enfoque que se expone en *d)* busca la construcción de agentes racionales cuyo estándar de desempeño no es la conducta humana, sino la racionalidad ideal.

El carácter racional del comportamiento del agente depende de cuatro factores que son:

- a) La medida con que se evalúa el grado de éxito del agente.
- b) De todo lo que hasta ese momento (previo a la elección de un curso de conducta) haya percibido el agente. A lo anterior se le denomina “historia perceptual” o “secuencia de percepciones”.
- c) Del conocimiento que posea el agente acerca del medio o ambiente en que está inmerso.
- d) De las acciones que el agente pueda emprender.

Un agente racional según Russell se define de la siguiente manera:

En todos los casos de posibles secuencias de percepciones, un agente racional *deberá* emprender todas aquellas acciones que favorezcan obtener el máximo de su medida de rendimiento, basándose en las evidencias aportadas por la secuencia de percepciones y en todo el conocimiento incorporado en tal agente.

La diferencia principal entre los enfoques que toman como estándar al comportamiento humano de los que toman el ideal de racionalidad consiste en que los primeros se conducen en un plano descriptivo, es decir, se plantean primeramente la descripción de la manera en que los humanos piensan y/o actúan, mientras que los segundos se conducen en un plano prescriptivo, esto significa que plantean la manera en que el agente debe pensar y/o actuar con base en ciertos supuestos independientemente de si coincida o no con el comportamiento que de hecho reportan los humanos.