

INTENCIONALIDAD E INTELIGENCIA ARTIFICIAL

Juan J. BOTERO C.

RESUMEN

Este artículo analiza la importancia que ha tenido la IA en la discusión filosófica, conceptúa sobre lo que es intencionalidad y valora su aplicabilidad en computadores y humanos, señalando las características de los procesos "mentales" que unos y otros pueden desarrollar, analizando las razones de fondo y opciones para responder los interrogantes asociados con la intencionalidad y la inteligencia.

LA INTELIGENCIA ARTIFICIAL EN LA DISCUSIÓN FILOSÓFICA

El interés de la filosofía por las investigaciones en inteligencia artificial (IA) podría medirse no por la enorme cantidad de trabajos filosóficos que esta nueva área ha suscitado sino, sobre todo, por su impacto sobre algunos sectores como la filosofía de la ciencia, la filosofía del lenguaje, la teoría del conocimiento ("epistemología"), sin hablar de la lógica. Pero tal vez el impacto mayor lo ha causado en el campo de la filosofía de la mente, desde donde podrían derivarse consecuencias interesantes para la ética, los metafísica, e indirectamente para otras disciplinas como la psicología, la antropología e inclusive los estudios jurídicos.

Este impacto puede resumirse en una idea que a primera vista puede aparecer perturbadora: la idea de que un dispositivo material pueda efectuar con éxito operaciones que pertenecen a capacidades mentales del tipo "inteligencia". En efecto, si una máquina resulta capaz de resolver problemas no mediante la aplicación ciega de recetas totalmente preestablecidas sino descubriendo por sí misma los métodos apropiados para su tratamiento, entonces se puede llegar a decir que esa máquina es inteligente. Ahora bien, como tradicionalmente se ha considerado a la inteligencia como una propiedad específica del ser humano, es lógico concluir que entre el ser humano y esta máquina no hay ninguna verdadera diferencia y que, por consiguiente, o la máquina es "humana" o el ser humano no es más que una máquina de este tipo, es decir, un sistema particularmente complejo de tratamiento de la información.

Ante la brutalidad de esta conclusión se ha sugerido que puede haber aquí un malentendido acerca del término "inteligencia", el cual tal vez no sea utilizado en un sentido idéntico cuando se trata de la máquina y cuando se trata del hombre. Es esta situación lo que ha hecho conveniente retomar y prolongar la reflexión filosófica acerca de la naturaleza del pensar dentro del nuevo contexto creado por las ciencias cognitivas en general y por la IA en particular. En medio de las discusiones a veces agrias generadas a este respecto, debe reconocerse que, desde la perspectiva de esta reflexión, el gran mérito de los trabajos que se adelantan dentro del programa de IA consiste en haberle dado una forma particularmente radical al problema clásico de las relaciones entre el pensar y el

cerebro, problema que a su vez es una especificación de la cuestión tradicional de la relación entre la mente (o el "alma") y el cuerpo.

Una manera de abordar la cuestión ha consistido en reflexionar sobre el uso que se hace en IA, tácita o explícitamente, de un concepto filosófico tradicional: el concepto de intencionalidad.

EL CONCEPTO DE INTENCIONALIDAD

Aunque el concepto de intencionalidad se remonta, como se sabe, por lo menos hasta la filosofía escolástica, el uso contemporáneo que se hace de él proviene de Franz Brentano y su intento de formular una "psicología desde un punto de vista empírico" a fines del siglo XIX. Brentano pensaba que para este propósito había que diferenciar claramente los fenómenos mentales de los fenómenos físicos y que esta diferencia podía establecerse del siguiente modo: todo fenómeno mental se caracteriza por estar referido a un contenido, por poseer lo que él llama una "objetividad inmanente" y que los escolásticos de la edad media llamaban "in-existencia intencional". Esta "in-existencia intencional" es característica exclusivamente de los fenómenos mentales y ningún fenómeno físico la posee. Por consiguiente, concluye Brentano, es posible definir los fenómenos mentales diciendo que son aquellos fenómenos que incluyen en su interior un "objeto intencional".

A partir de Brentano pueden seguirse dos vías para el tratamiento del tema de la intencionalidad: una es la de la fenomenología de E. Husserl, para la que la intencionalidad es la característica esencial de la conciencia y tiene una naturaleza trascendental. La otra, más recientemente explorada, se deriva del tratamiento "lingüístico" de la tesis brentaniana, en particular, del intento realizado por R. Chisholm de transformarla en una tesis a propósito de las propiedades lógicas de las oraciones referidas a fenómenos mentales. Según Chisholm, lo que caracteriza a los enunciados "intencionales" (enunciados acerca de fenómenos mentales) es que son "intensionales" (no extensionales).

Aunque por esta vía se generaron numerosas y a veces interesantes discusiones lógicas, lo importante de ella para nosotros no son estas discusiones sino el hecho de haber introducido dentro del ámbito de la llamada "filosofía analítica" el tema de la intencionalidad ligado al de las "actitudes proposicionales" (enunciados que refieren estados mentales: creencias, deseos, etc.). Es en el contexto de la discusión acerca de la naturaleza de las actitudes proposicionales que el concepto de intencionalidad ha jugado un papel interesante en el cognitivismo y en la reflexión filosófica acerca de la IA en particular. Y es del sentido crítico de esta reflexión de lo que me voy a ocupar en lo que sigue, utilizando para ello la crítica que hace John R. Searle a la IA valiéndose precisamente de este concepto.

La crítica de Searle puede resumirse de este modo: el Cognitivismo y la IA "Fuerte" están errados al comparar al pensamiento humano y a la mente en general con los computadores por una razón muy simple: ni los computadores ni los programas pueden ser "mentes" porque, a diferencia del pensar y de la mente humanas, no poseen intencionalidad.

INTENCIONALIDAD, COMPUTADORES Y HUMANOS

Searle se pregunta por la significación psico y filosófica de los esfuerzos por simular computacionalmente las capacidades cognitivas humanas. Y distingue dos tendencias: una "débil", para quien estas simulaciones son una herramienta útil que nos permite formular y someter a prueba hipótesis de un modo riguroso y preciso. Pero hay también una tendencia "fuerte" en la IA según la cual el computador no es sólo una herramienta. Apropiadamente programado, un computador es realmente una mente, en el sentido de que si se le dan los programas correctos se puede afirmar que comprende y tiene lo que se llaman "estados mentales". Desde el punto de vista de la psicología, lo que esta tendencia sostendría es que los programas no son sólo herramientas útiles para entender los procesos cognitivos sino que ellos son las explicaciones psicológicas. Es contra esta tendencia que dirige su crítica.

En un comienzo la crítica se dirige a los trabajos de Schank y sus colegas de Yale sobre los Guiones, pero ella tiene que ser general, y es mejor entenderla de una vez así.

El concepto que tenemos del computador digital implica forzosamente que sus operaciones puedan especificarse de manera puramente formal: se expresan las etapas de la operación de cómputo en términos de símbolos abstractos; una regla específicamente computacional determina cómo se "tratan" esos símbolos (recordar la máquina de Turing). Ahora bien: los símbolos no tienen significación, no tienen contenido semántico; por consiguiente, no tratan de nada. Deben especificarse únicamente en términos de estructura formal o simbólica. Esta característica es precisamente la que hace tan poderosos a los computadores digitales.

Por otra parte, el funcionamiento de la mente humana no se resume en procesos formales o sintácticos. Por definición, los estados mentales tienen un contenido, son Intencionales. Y Searle especifica su lectura de la tesis brentiana del siguiente modo: decir que un estado mental es Intencional es decir que él es una representación de sus condiciones de satisfacción. Con otras palabras, los estados mentales Intencionales son siempre sobre algo, acerca de algo. La mente, además de sintaxis, tiene semántica, en el sentido de que, además de su estructura formal, tiene un contenido, un sentido.

EL CUARTO CHINO

Para ilustrar su crítica, Searle pone un ejemplo que se puede generalizar así: supongamos que un grupo de hábiles programadores hayan escrito un programa que permita a un computador simular la comprensión del idioma chino: pasa el Test de Turing. ¿Se puede decir que el computador comprende el chino? Imaginemos ahora una situación similar, pero que nos involucre a nosotros. Searle lo llama "el experimento del cuarto chino".

Supongamos que a una persona que no sabe nada del idioma chino se la encierra en un cuarto que contiene numerosos cestos repletos de símbolos chinos. También se le da a esta persona un manual en español para que haga corresponder símbolos chinos con otros símbolos chinos según ciertas reglas. Estas reglas identifican a los símbolos únicamente por la forma que tienen y no requieren que se los comprenda. Podrían ser de este tipo: "tome un signo que tiene la forma tal del cesto número uno y colóquelo enseguida de un signo de

forma tal otra del cesto número dos", o algo similar. Ahora imaginemos que quienes están fuera del cuarto y que comprenden el chino introducen trozos de papel con símbolos, y que en respuesta la persona del interior los manipula según las reglas del manual y luego envía al exterior otros trozos de papel con símbolos así manipulados. Pues bien: el manual de instrucciones (las reglas) sería el "programa del computador". Quienes lo escribieron serían los "programadores". Y la persona encerrada en el interior del cuarto sería el "computador". Los cestos repletos de símbolos serían la "base de datos", los trozos de papel introducidos serían las "preguntas" y los que la persona envía al exterior las "respuestas". Si el manual está bien escrito de manera que las "respuestas" a las "preguntas" no se puedan distinguir de las que daría un hablante nativo del chino, se diría que la persona que se encuentra en el interior "satisface el test de Turing", en este caso, para la comprensión del idioma chino.

Ahora bien, es claro que quien está en el cuarto chino no entiende ni jota de chino, y no hay forma de que lo entienda utilizando ese sistema pues no hay forma de que aprenda el significado de ninguno de los símbolos. La persona encerrada en el cuarto chino hace lo mismo que un computador: manipula símbolos, pero sin acordarles ninguna significación. La moraleja es ésta: si esta persona no comprende el chino basándose únicamente en la aplicación de un programa para entender el chino (el manual y sus reglas), entonces un computador tampoco puede hacerlo cuando trabaja sobre esta misma base. Y lo que vale en este caso para el idioma chino vale en general para otras formas de comprensión y de conocimiento. La mera manipulación de símbolos (en realidad son meros garabatos puesto que no tienen sentido para quien los manipula) no es por sí misma suficiente garantía para el conocimiento, la percepción, el pensar, etc. Piénsese en la diferencia que hay entre esa situación y una en la cual las preguntas son en español y las respuestas también deben ser en español: a diferencia del primer caso, en el segundo nosotros comprendemos las preguntas, porque están formuladas en símbolos cuya significación conocemos, es decir, que tienen sentido para nosotros. En el primer caso, sólo manipulamos símbolos según un programa (reglas), sin acordarles ninguna significación.

ESTADOS MENTALES E INTENCIONALIDAD

Ahora quisiera reflexionar brevemente sobre el sentido y los alcances de este tipo de críticas.

¿Por qué no tienen intencionalidad los computadores o sus programas?

Searle dice que porque son exclusivamente formales. Construir una máquina cuyo programa estuviera completamente "cableado", y que además tuviera interacciones con el medio, resolvería el problema? Según Searle no, pues la intencionalidad depende de la bioquímica de sus orígenes. Es decir: ella es producida en el SNC por los poderes causales de este mismo SNC, de modo que sólo podrían tenerla criaturas que poseyeran un SNC similar al nuestro en cuanto a sus poderes causales. Este es un asunto empírico.

Sin entrar a esta "ontología", aún es posible plantearse la pregunta:
¿Cómo adquieren nuestros estados mentales la intencionalidad?

No se trata de una pregunta necesariamente empírica. Puede entenderse como una pregunta acerca de la naturaleza del conocimiento humano, sus fuentes, sus alcances. En la historia de la Filosofía ha habido diversas maneras de responderla.

- a) Una es la opción trascendental, como la de Kant o Husserl: hay estructuras de la subjetividad que dan cuenta de este fenómeno. En Kant tenemos las formas de la sensibilidad y del entendimiento, con su esquematismo y su estructura de reglas. En Husserl, las estructuras de la conciencia son en sí mismas Intencionales, es decir, poseen un contenido, lo cual quiere decir que no son "puras formas", como en Kant. En su descripción fenomenológica de estas estructuras Intencionales Husserl llega a formulaciones muy cercanas a las de la IA, lo cual ha hecho que algún autor (Dreyfus) llegue a catalogar a la fenomenología como un antecedente de la IA: la descripción de los niveles intencionales en la constitución del objeto intencional, por ejemplo, es muy cercana a los "marcos" de Minsky.

Pero el propio Husserl llega a un punto crítico en la fenomenología al reconocer que la descripción fenomenológica es una "tarea infinita". Y no es difícil ver por qué.

Como los programas de representación del conocimiento en IA, la descripción Intencional de Husserl se encuentra con que la determinación de cualquier concepto lo lleva a otra multitud de conceptos en varios niveles, cada uno de los cuales a su vez hace necesario ampliar aún más la descripción. Cada contenido Intencional se determina así en medio de una red de contenidos que no puede ser fijada completamente. El resultado es lo que se conoce como "explosión combinatoria".

Pero además, Husserl se encuentra con que los estados mentales Intencionales constituyen su contenido siempre sobre la base de otro tipo de intencionalidad, que él llama "operante": esta intencionalidad no es representacional como la otra, sino precisamente "operante", es decir, constitutiva de sentidos de un modo no predicativo. Son lo que él llama experiencias del tiempo, del espacio relativo a los propios movimientos (cinestésias) y de asociaciones motivadas. Lo que cuenta, pues, en este nivel básico y fundamental de la intencionalidad, es la experiencia de la vida cotidiana, la cual, al no ser representacional, no puede expresarse como "conocimientos". Al llegar a este punto, la fenomenología alcanza su límite y permite ver la otra opción.

- b) Podemos llamar "opción existencial" a la segunda vía para dar cuenta de la intencionalidad. Llegado al punto en donde la especificación del sentido, es decir, de los contenidos intencionales, ya no consiste en establecer redes de creencias, de expectativas (todas ellas representaciones), y de reglas, hay que saber por qué se llega a ese límite. ¿Se trata realmente de una "tarea infinita", o más bien de una "tarea equivocada"? Es posible que el "sentido", es decir, la intencionalidad no sea un asunto exclusivamente "mental". Se puede pensar que sea un asunto que involucre a la totalidad del ser humano, con su mundo de objetos, de personas, de instituciones, etc.

El contexto que permite establecer los contenidos Intencionales de los estados mentales, entonces, no estaría dado por otros conocimientos, creencias, etc., sino por las prácticas sociales y culturales, los artefactos, etc., así como por las circunstancias

biológicas, históricas y físicas en las cuales se vive, se percibe, se habla, se ama, se desea, etc. A todo este contexto podemos darle el nombre de "situación". Y podemos decir entonces que lo que constituye y estructura la intencionalidad son las "situaciones" reales en las que vive el ser humano. Y estas "situaciones" no son estructuras de creencias o de representaciones de ningún tipo.

Esta segunda opción es un desafío filosófico al enfoque cognitivista para el cual el "mundo del sentido común" debe ser, de un modo u otro, un "sistema de creencias". Pero desde la propia IA se ha vuelto cada vez más difícil sostener ese enfoque, dada la estimación del tamaño y complejidad crecientes del sistema de creencias requerido. También ha sido muy importante el poco progreso obtenido en el tratamiento del "problema del marco". Como se sabe, las estimaciones acerca del número de creencias involucradas en el conocimiento cotidiano han crecido desde la estimación de Minsky en 1968 ("cien mil elementos de conocimientos necesitará una máquina para comportarse con razonable sensibilidad en una situación ordinaria"- Sam Inf. Procesan) hasta la de Dennett en 1984 ("sabemos trillones de cosas"- Cognitive Wheels: The Frame Problem of AI).

COGNITIVISMO, INTELIGENCIA E INTELIGENCIA ARTIFICIAL

El cognitivismo como "punto de vista" no está necesariamente casado con una metafísica de hechos y reglas. Todo lo que el punto de vista cognitivo implica en última instancia es la tesis de que el sentido del mundo y de los fenómenos humanos se constituye a partir del ser humano. Las estructuras cognitivas pueden entenderse entonces en un sentido más amplio que el tradicional de la "teoría del conocimiento".

Justamente, uno de los logros de este punto de vista puede ser el abandono de toda "teoría del conocimiento", que es lo que caracteriza también el paso de la "opción trascendental" a lo que hemos llamado, a falta de un término mejor, "opción existencial". Este cambio se basa en la profundización heideggeriana de la fenomenología husserliana y consiste esencialmente en cuestionar el privilegio que Husserl le otorga a la teoría del conocimiento y en mostrar que las diversas modalidades del "yo" intencional se fundan en el ser-en-el-mundo. Dicho de otro modo, toda "cognición", y en particular todas las modalidades de la comprensión cotidiana, como por ejemplo la percepción de los objetos que nos rodean, o el ejercicio cotidiano del lenguaje, presupone esa apertura a un mundo para un ente que está "situado" en él y que se "proyecta" en él hacia posibilidades indeclinables, según la particular terminología heideggeriana. La "inteligencia" de la IA no es un dato básico pues ella presupone este ser-en-el-mundo. Para decirlo más radicalmente aún: la inteligencia es ante todo la autocomprensión del ser-en-el-mundo, pues lo que caracteriza a este ente que es el ser humano es precisamente la comprensión, que consiste en estar consagrado (o condenado, según se le quiera mirar) al sentido y en saberse tal.

El programa de la IA podría encontrar así sus límites, que no serían otros que los límites de la mathesis cartesiana, fundamento del pensar mecanicista y calculista en general. El intento por hacer entrar dentro de esta mathesis a la intencionalidad convirtiéndola en un sistema de hechos y reglas de procedimiento no debería entonces perder de vista las limitaciones de su aplicación.

Tal vez encontremos aquí la explicación del carácter perturbador de la idea básica de la IA que mencionaba al principio. Y también la confirmación de la sospecha de que existe un malentendido esencial ligado al uso del término "inteligencia" aplicado a seres humanos y a máquinas. Este malentendido podría provenir de la extensión más allá de sus límites de los propósitos de la IA. Y como tal no sería más que la manifestación del malentendido general que ha caracterizado a toda la modernidad.