

## Una aproximación a la libertad desde la IA

¿Cómo funciona la IA?

Partiendo de una definición sencilla de la IA, como la de aquellas máquinas capaces de imitar determinadas capacidades cognitivas humanas, la primera pregunta que nos podemos hacer es qué es lo que nos sorprende realmente para calificar una máquina como inteligente.

Esta pregunta es especialmente necesaria en un contexto, como el actual, en el que los ordenadores están absolutamente generalizados en nuestra sociedad y pueden realizar complejas operaciones en tiempos infinitamente pequeños respecto al tiempo equivalente que necesitaría una persona.

Las generaciones más jóvenes, además, se sorprenden mucho menos con determinados avances en la IA en la medida en que su punto de referencia es ya muy exigente.

Sin embargo, seguramente sí podemos afirmar que la capacidad desarrollada por algunos sistemas de IA para identificar enfermedades con precisión ante el análisis de determinadas imágenes, la capacidad de un sistema de mantener diálogos fluidos con una persona y responder de forma natural a sus preguntas o incluso de crear arte, nos acercan al concepto de inteligencia artificial que estamos buscando.

Desde esta aproximación, lo que es relevante es identificar el proceso subyacente a esas actividades que calificamos como inteligentes. Para ello, podemos descomponerlo en sus componentes principales:

- Los datos, como elemento imprescindible para el proceso de aprendizaje (es decir, la capacidad que permite que una máquina modifique sus respuestas en función de los datos que va recibiendo, si de alguna manera esto se puede llamar experiencia).
- El proceso de decisión (algoritmo), que, a su vez, requiere de diferentes elementos:
  - Los datos que recibe para el caso concreto.
  - Unos principios generales de definición de funcionamiento del algoritmo.
  - Un objetivo que intenta lograr, mediante la optimización de las posibles alternativas

El objetivo del algoritmo de decisión como núcleo del proceso

La clave del proceso está en el objetivo. Este es el elemento alrededor del cual gira el proceso. El objetivo está definido por el constructor y es independiente de los principios de funcionamiento del algoritmo. El arte del diseño de un algoritmo es combinar unos principios de funcionamiento que sean los más adecuados para la consecución del objetivo. Incluso se podría permitir que la máquina probara diferentes principios o estructuras de algoritmos para que encuentre aquella que le permite optimizar en mayor medida las probabilidades de las soluciones ofrecidas.

No obstante, en un caso u otro, lo que la máquina nunca puede hacer es modificar el objetivo. No tiene esa capacidad.

Para dotarle de esa posibilidad se podrían adoptar dos opciones: incluir un proceso de aleatoriedad que permita variaciones del objetivo o bien introducir un objetivo de orden superior sobre un conjunto de algoritmos, de forma que exista una jerarquía que modifique los

objetivos de los algoritmos inferiores en función de cómo estos van contribuyendo al objetivo superior.

En ambos casos, lo que podemos deducir es que, por construcción, los algoritmos y los procesos basados en inteligencia artificial no pueden tener libertad (entendido como la posibilidad de elegir por si mismos sus objetivos y, en función de ellos, modificar sus procesos de decisión).

Por tanto, una máquina inteligente siempre estará determinada a cumplir el objetivo definido, salvo que otra máquina le pueda imponer una modificación del objetivo o que introduzcamos una variable aleatoria. Pero en ninguno de ambos casos la máquina inteligente sería libre (si, además, pudiera llegar a tener consciencia de esa libertad).

En el caso de los humanos, quizás podamos llegar a decir que esa libertad, esa posibilidad de cambiar de objetivo según las circunstancias, está guiada por meros efectos bioquímicos que se producen en nuestro cuerpo en función de las circunstancias del momento. Bajo esa visión, sería posible también introducir un mecanismo semejante en nuestra máquina inteligente.

No obstante, nuestra experiencia personal seguramente nos lleva a pensar que nuestra capacidad de cambiar de objetivo no esta sólo determinada por la bioquímica de las circunstancias, sino que nuestra capacidad cognitiva puede introducir modificaciones en el propio proceso que, aunque tengan igualmente un sustrato biológico, su origen está en la voluntad del sujeto.

Traslación a un caso real

Tratemos de visualizar esto en un caso concreto. Imaginemos una persona que tiene que decidir sobre si concede o no determinada petición a otra.

El decisor, inicialmente, pediría al sujeto que le proporcionara los datos que habitualmente usa para tomar la decisión.

Hasta aquí habría poca diferencia con un proceso basado en inteligencia artificial. No obstante, el decisor habrá recibido y percibido, consciente o inconscientemente, infinidad de datos que la máquina no habrá procesado.

En unos casos la inteligencia artificial no tendrá los sensores que le permitan recoger esa información (olores, entonación, forma de andar del sujeto, sonrisa, etc.).

Aun en el caso de que pudiera llegar a recoger todos esos datos, la máquina los despreciaría en la mayor parte de los casos, o bien porque no tiene ese dato en el resto de los casos y, por tanto, no tiene elementos de comparación para poder valorar el dato o porque teniendo esos datos en otros casos, considera que es un dato no concluyente para el objetivo perseguido, descartándolo del proceso posterior.

Una vez recabados los datos, el decisor usará el algoritmo de decisión correspondiente para tomar una decisión, de aprobación o denegación de la petición.

En el caso del decisor humano, es posible que antes de denegar la petición, haga alguna pregunta o consideración adicional para intentar ayudar al sujeto. La intensidad de este proceso puede verse afectada por las circunstancias personales y coyunturales del decisor. Este argumento es usado para defender la mayor objetividad del proceso basado en la inteligencia artificial.

En el caso de la máquina, podría incluirse en su proceso de decisión una rutina semejante, de forma que intente identificar variables que harían que el proceso acabara en una decisión favorable y busque, en un diálogo con el sujeto, si dichas variables se cumplen en su caso.

Tras esta secuencia de contraste de preguntas y respuestas, el decisor, ya sea humano o una inteligencia artificial, tomará la decisión. Si la decisión es contraria a los intereses del sujeto, posiblemente este reaccionará intentando añadir más información o reiterando la ya existente para que su petición sea reconsiderada.

Aquí, volveríamos al estadio anterior. El decisor humano, dependiendo de diversas circunstancias, podrá hacer un esfuerzo mayor o menor por evaluar de nuevo la situación. En el caso de la inteligencia artificial, se podría simular nuevamente esa revisión mediante un nuevo despliegue de alternativas para que el sujeto pueda aportar algún elemento adicional que cambien la decisión.

Pero lo que nunca hará la inteligencia artificial es modificar su objetivo, que es maximizar la posibilidad de acierto en la resolución adecuada de la cuestión.

Sin embargo, en el caso del decisor humano, podría darse dos escenarios.

Un primer escenario sería decidir, en un momento dado del proceso, darle mayor o menor peso a una circunstancia determinada y, en función de ello, tomar una decisión diferente. En este caso se podría decir que está alterando las reglas de decisión sin fundamento, por lo que estaría resultando en una decisión no óptima desde el punto de vista más objetivo de la inteligencia artificial. No obstante, algo que podría suceder es que el decisor humano tome en consideración una circunstancia nueva, sobre la que no existen antecedentes, pero que le otorgue el suficiente valor como para modificar su decisión. La inteligencia artificial difícilmente puede evaluar algo novedoso si no tiene patrones de referencia previos.

Y existe un segundo escenario también probable. Este viene dado porque, valorando las circunstancias del caso, el decisor humano considera que hay un objetivo de orden superior que deben prevalecer y, por tanto, toma la decisión valorando el resultado óptimo para ese principio superior, aunque sea contradictorio con el objetivo inicial. Incluso, en determinadas circunstancias, puede llegar a tomar una decisión que favorece al sujeto, aunque le perjudique a el mismo o le ponga en riesgo, incluso sin ningún beneficio para él o para sus congéneres.

Estas posibilidades son lo que llamamos libertad. Estas posibilidades son las que nunca podrá gestionar una inteligencia artificial.

Podrá utilizar un conjunto jerárquicamente organizado de principios, pero no tomará decisiones autónomas e incluso en algunos casos contradictorias, sobre la estructura de dichos principios. Sus decisiones están, de alguna forma, predeterminadas, aunque puedan permanecer ocultas a nuestro entendimiento tras una vasta maraña de algoritmos y procesos de aprendizaje. Pero todos ellos encaminados a un fin último que es tratado en forma probabilística y cuya decisión será siempre la que genere mayor probabilidad, minimizando el riesgo de error. No hay espacio para la locura.