

El término “inteligencia artificial” (IA), que se adoptó en la Conferencia de Dartmouth de 1956 a instancias del científico de la computación John McCarthy, no es el más acertado, pero ha llegado para quedarse, por su simplicidad, y porque a nadie se le ha ocurrido una alternativa. Se refiere a un modelo creado para resolver un problema específico o prestar un servicio concreto.

Partamos de un ejercicio mental. Si, efectivamente, como piensan algunos astrofísicos, no se puede ir a mayor velocidad que la luz, los viajes interestelares y no digamos los intergalácticos, serán muy difíciles para los humanos, porque no tendrían capacidad vital para cruzar espacios de miles o millones de años luz. Cabe pensar lo mismo de una civilización extraterrestre que nos descubriera, aunque las naves que llegaran a la Tierra fueran capaces de transportar alguna vida biológica (como esporas o bacterias) que pudiera desarrollarse después en formas más complejas. Hay que estar dispuestos a recibir a máquinas inteligentes, que se habrían perfeccionado a sí mismas en el tiempo del larguísimo viaje, y que probablemente discurrirían de otras formas distintas a las humanas. U otras formas, como la comunicación a través de sofones (protones equipados de inteligencia) que Liu Cixin ideó en su magnífica trilogía *El problema de los tres cuerpos*.

Por “inteligencia artificial general” (IAG), un nivel mucho más alto en comparación a la IA, se suele entender un programa o *software* capaz de aprender y la mayoría de las tareas intelectuales que hacen los seres humanos, incluido el propio desarrollo de IA. Estamos aún lejos de eso. Y quizás no llegue nunca, sino otra cosa. Noam Chomsky y otros colegas desechan “ese momento largamente profetizado en el que las mentes mecánicas superarán a los cerebros humanos no sólo cuantitativamente en cuanto a velocidad de procesamiento y tamaño de la memoria, sino también cualitativamente en cuanto a perspicacia intelectual, creatividad artística, crítica y cualquier otra facultad distintivamente humana. «Ese día», dicen, «puede llegar, pero aún no ha amanecido».

Un problema, entre otros muchos, de este tipo de visión es que la IA suele estar referida a la inteligencia humana. De hecho, el cerebro humano se puede considerar una máquina de Turing universal, aunque más lento, pero muy eficiente en términos energéticos. Es una inteligencia con una memoria limitada y cambiante como indica la neurociencia e intuyó Marcel Proust, más útil para pensar el futuro, para planear, que para recordar el pasado. Tiene muchas dimensiones. Dentro de la inteligencia humana hay varios tipos de inteligencia. La inteligencia artificial se suele definir como “la simulación de la inteligencia humana por parte de las máquinas, especialmente sistemas informáticos”. Es, además, en sus principios, un producto de la inteligencia humana, condición de la que la IA puede escapar. Esa definición fue propuesta por McCarthy, pensando en procesos que incluyen el aprendizaje (la adquisición de información y reglas para el uso de la información), el razonamiento (usando las reglas para llegar a conclusiones aproximadas o definitivas) y la autocorrección.

Desde entonces, ha sido ampliamente adoptada y es la base de muchos desarrollos actuales en IA que intentan emular para superarlo el funcionamiento del cerebro humano, del que queda mucho, muchísimo, por conocer, tanto su *software*, como su *hardware*. Pero no cabe olvidar que la naturaleza nos ha surtido con otros tipos de inteligencia, naturales, como la de muy diversos animales, la colectiva de las hormigas u de otros seres biológicos mucho más pequeños, o la que se refleja en las migraciones estacionales de algunas aves. Puede que lleguemos a varios tipos de inteligencia o capacidades de máquinas más allá del mero procesamiento de información.

En la inteligencia humana, emoción y razón van ligadas

Quizás el término “conocimiento” hubiera sido más adecuado, sobre todo referido al “explícito”, pues el “tácito”, muy humano, es muy difícil de conseguir de forma artificial. La construcción de sentido, los valores, los juicios, la interacción social y la cuestión de la empatía, para crear y generar conocimiento, entre otros, son necesarios. Y no porque implique “conciencia”, pues seguimos sin saber realmente qué es, pese a años de investigación. Desde Freud sabemos que también tenemos un inconsciente, y desde mucho antes que gran parte de nosotros, de nuestros cuerpos, funcionan de forma inconsciente para nosotros. Para Freud, indica el filósofo esloveno Slavoj Žižek, «la decisión también es anterior a la conciencia; sin embargo, no se trata de un proceso puramente objetivo, sino de una decisión inconsciente». En la inteligencia humana, emoción y razón van ligadas. Se plantea también la cuestión de si es posible la inteligencia sin un cuerpo. La mente no se entiende sin el cuerpo, por mucho que le pese a René Descartes y a algunos “inmortalistas” que creen que la mente humana se puede trasladar a un ordenador. Tampoco la IA es, al menos de momento, “sintiente” (término que utilizaba el filósofo/teólogo Xabier Zubiri, y no “sensible” como algunos traducen, mal, el *sentient*), es decir, con capacidad de poseer sentimientos. Aunque sí va teniendo la capacidad de detectar algunos de nuestros sentimientos o estados de ánimo. Algo importante, por ejemplo, para los cuidados a cargo de máquinas.

Para el humano, al menos en la cultura occidental, hay esencialmente dos formas de conocimiento: la ciencia y el arte. Lo artificial ha sido durante mucho tiempo lo fabricado por los humanos. Con la IA puede nacer una nueva forma de conocimiento que nos permita ir más allá de la lógica o razonamiento de la inducción (las premisas apoyan la conclusión, pero a menudo no la garantizan, pues el resultado se basa en datos siempre incompletos por muy masivos que sean), de la deducción (la conclusión sigue a las premisas) o de la abducción (buscar con una conjetura la explicación mejor o más probable). No entramos aquí en el tema central de la intuición. Aunque se están investigando otras lógicas para las máquinas más deductivas, la llamada IA actual es fundamentalmente estadística, inductiva. Busca pautas en ingentes cantidades de datos que un humano por sí solo no abarcaría. Lejos de lo que plantea Daniel Kahneman¹, para el cual los humanos tenemos el Sistema 1 (rápido, instintivo, emocional y subconsciente) y el Sistema 2 (más lento, deliberativo, lógico y consciente). Ambos sistemas interactúan en la toma de decisiones. Quizás la IA no llegue nunca a eso, pero puede abrir un nuevo Sistema 3 aún desconocido.

Los LLMs (*large language models*, o modelos amplios de lenguaje), como el chat GPT y otras IA generativas basadas en lógicas esencialmente bayesianas², seleccionan en el texto o información que se le somete pautas de lenguaje, de temas, y devuelve la respuesta en forma de textos, imágenes, música, códigos y fórmulas, bastante bien escritos, ya que los ha tomado de alguna parte, aunque aún de poca fiabilidad. Los LLMs parecen aprender y usar representaciones del mundo externo. Pero es solo una apariencia, basada en la tecnología del *machine learning*. Es imitación, cuando no directamente copia, de textos u otros productos humanos, en una dimensión enorme. Para los puristas, que han perdido esta batalla semántica, no es propiamente inteligencia, porque no se parece a la humana, aunque se usen términos para otros avances como las “redes neuronales” (también utilizadas en este contexto) inspiradas en el cerebro humano. Pero no hay quien le quite ya el apelativo, menos aún cuando empieza a generar, y no solo comprobar, hipótesis.

Si se pregunta a Chat GPT si dice ser una IA: “Soy un modelo lingüístico de inteligencia artificial (IA) desarrollado por OpenAI. He sido entrenado en un gran conjunto de datos de texto de Internet utilizando algoritmos de aprendizaje profundo para generar respuestas similares a las humanas a preguntas basadas en texto. Mi objetivo es ofrecer respuestas útiles e informativas a usuarios como tú”. El famoso test de Turing para saber si estábamos hablando con una máquina o con un humano ha quedado obsoleto. Ahora bien, estos nuevos *chats* nos enseñan algo sobre nosotros mismos. Como indica el filósofo Antonio Diéguez: “una de las

lecciones a sacar del ChatGPT y similares es cuánto hay de automatizable y previsible en lo que decimos y escribimos”. Žižek va más lejos al hablar de “idiotez artificial”. Refiriéndose a la generativa de estos *chatboxes*, Robert Shrimmsley³ considera que “la humanidad ha construido la IA en su propia imagen: una de chanchullos, trampas y chorradas para resolver un problema”. Aunque no es seguro que la IA, a medida que se vuelva más autónoma, si seguirá esta pauta, la de la propia imagen.

La manera de no quedarnos atrás los humanos puede ser conectarnos, de diversas formas, a esas máquinas

Ahora bien, los LLMs son algo más que *machine learning* (cuyo centro es descripción y predicción, establecimiento de correlaciones, pero no de causalidades), pero algo bastante menos que inteligencia. Muchos expertos que han desarrollado esta IA generativa se dicen incapaces de interpretar cómo funcionan por dentro. Y estamos tan sólo en el principio. Un nuevo estudio de Colin West⁴, que compara resultados de GPT 3.5 y GPT 4 en problemas de Física de bachillerato muestra que en unos pocos meses, esta IA ha pasado del nivel del 39% más bajo al de del 96% de los alumnos, con lo que aprobaría (eso sí, sin metáforas). ¿Y en cinco años? ¿Y en diez? ¿Y cuando estos sistemas se encadenen o se incluyan en amplios grupos de programas, en una IA colectiva, además de la inteligencia colectiva humana en sí?

En su libro sobre la IA *The Age of AI: And Our Human Future*, Henry Kissinger, Eric Schmidt⁵ y Daniel Huttenlocher⁶, ya señalaban que “los humanos están creando y proliferando formas no humanas de lógica con un alcance y una agudeza que, al menos en los entornos discretos para los que fueron diseñados, pueden superar la nuestra”. Citan el caso del ordenador AlphaZero (de DeepMind de Google) que en 2017 aprendiendo solo -no con miles o millones de jugadas de humanos, sino jugando contra sí mismo- generó nuevas formas de ajedrez que los grandes maestros no habían visto nunca antes y de las que estos aprenden. David Silver, investigador científico principal de DeepMind, insistió en 2016 en que no podía explicar cómo AlphaGo -otro programa diseñado por él- logró desarrollar su estrategia ganadora: «Lo descubrió por sí mismo a través de su propio proceso de introspección y análisis». Puede indicar los albores de una nueva forma de conocimiento explícito codificable, aunque, al menos en el caso del ajedrez, dentro de marcos diseñados por los humanos. Por ahora. Pese a toda la IA capaz de ganar a los mejores maestros, reconforta el interés por el ajedrez entre humanos, que se ha plasmado en el último campeonato del mundo que ha ganado el chino Ding Liren, frente al ruso Ian Nepomniachtchi. La llegada de la IA, dicen estos autores, “nos obliga a enfrentarnos a si existe una forma de lógica que los humanos no han alcanzado o no pueden alcanzar, explorando aspectos de la realidad que nunca hemos conocido ni podremos conocer directamente”. Nuevas formas de inteligencia.

La manera de no quedarnos atrás los humanos puede ser conectarnos, de diversas formas, a esas máquinas. De hecho, ya hemos empezado a hacerlo a través de esa extensión de nuestro cerebro en que se han convertido los *smartphones*. Piénsese, o imagínese, lo que puede dar de sí que los humanos lleguemos a integrar en nuestros cerebros otras formas de conocer. Nos cambiará. No digamos ya mezclada con la manipulación genética y biológica. La combinación entre más capacidad biológica humana y tecnológica de las máquinas, abrirá la puerta a una nueva forma de inteligencia: la trans-inteligencia. Como dice Žižek (Hegel y el cerebro conectado, 2023) “promete una nueva experiencia subjetiva de inmersión en el espacio de la mente colectiva”. Sea como sea a la inteligencia artificial hay que echarle imaginación, que es algo muy humano.