

El oxímoron de la inteligencia artificial

Los atributos principales de la inteligencia no están presentes en las actuales “inteligencias artificiales”, que son incapaces de pensar, abstraer, inferir o comprender. El proceso de razonar, o la capacidad de comprender el significado de lo que se dice o el sentido de lo que ocurre, está fuera de su alcance actualmente.

Dentro de las figuras lógicas literarias, el oxímoron implica contradicción. En efecto: cuando tantas veces hablamos de los “muertos vivientes” (o zombis) estamos estableciendo un concepto totalmente contradictorio. Pues bien: cuando leemos todos los días que una “inteligencia artificial” (IA) ha realizado un nuevo prodigio, estamos introduciendo una evidente contradicción, porque dichas “inteligencias” no son inteligentes.

Atributos esenciales para definir la inteligencia

El concepto de inteligencia no está definido de forma unívoca, a pesar de que desde hace milenios gran número de especialistas en filosofía, fisiología, matemáticas, física, ciencias de la computación, neurociencia y psicología han tratado de explicar y acotar sus aspectos esenciales. En cualquiera de los casos, algunos de los atributos principales de la inteligencia que reconocemos bien, como la consciencia o la capacidad de pensar, abstraer, inferir o comprender, nos permiten comparar en parte la inteligencia natural de las personas con la inteligencia artificial de las máquinas.

Debemos clarificar que las modernas IA son básicamente herramientas muy útiles para facilitarnos la realización de algunas de nuestras actividades, pero no son inteligencias, dado que incluso las características más básicas de la inteligencia que hemos enumerado no las incorporan en absoluto. De los atributos anteriormente citados, quizás el más fácilmente implementable en una máquina mediante un algoritmo es la inferencia, puesto que la lógica es una de las sólidas ramas de las matemáticas, y por tanto es programable de múltiples maneras en un computador.

El proceso de razonar

Inferir es extraer un juicio o una conclusión a partir de un proceso de razonamiento lógico. La inferencia es algo que todas las personas efectuamos constantemente en nuestra vida cotidiana, y podemos preguntarnos si las IA también infieren de forma similar.

ChatGPT tiene muchos admiradores actualmente, pero no es una inteligencia y no comprende nada de

Lo que dice

Para entender mejor este concepto, podemos recurrir al famoso detective literario Sherlock Holmes. En la primera novela de esta saga de Arthur Conan Doyle (*Estudio en Escarlata* en 1887), el Doctor Watson describe, en el segundo capítulo “la ciencia de la deducción” del protagonista. Realmente el razonamiento deductivo en ciencia parte de una hipótesis teórica, que sigue de observaciones o experimentos y concluye con confirmación (o no) de la hipótesis inicial. Normalmente Sherlock Holmes utiliza el razonamiento inductivo, partiendo de observaciones que le permiten establecer una hipótesis y una teoría de los hechos que realmente ocurrieron.

Si vamos un poco más allá, en la segunda novela de la saga (*El signo de los cuatro* en 1890), en el sexto capítulo Sherlock Holmes hace una demostración y espeta a Watson “¿Cuántas veces le he dicho que una vez eliminado lo imposible, lo que queda debe ser la verdad, por improbable que parezca?”. Este tipo de inferencia fue denominada «razonamiento abductivo», por el científico y filósofo Charles S. Peirce en 1878. Este razonamiento, incorporado a la inteligencia natural humana, implica cuestionarnos constantemente por qué ocurre algo y formular hipótesis sobre causas plausibles para ello. Es muy llamativo que ninguna de las IA actuales ni siquiera se acerquen a realizar este tipo de razonamiento, tan común entre nosotros, y a lo sumo se utilizan grandes bases de datos y un cálculo estadístico para intentar remedar una inteligencia, pero sin ningún tipo de razonamiento lógico como los que hemos descrito.

Comprender el significado de lo que se dice o el sentido de lo que ocurre

Ingredientes más complejos de la inteligencia, como nuestra capacidad de comprender, están completamente fuera de juego en las IA actuales. A pesar de todo, muchas compañías nos anuncian frecuentemente “nuevas inteligencias artificiales generativas”, capaces de “crear” contenidos nuevos y originales, sin intervención humana.

Realmente ninguna de estas herramientas son inteligencias y ninguna de ellas crean algo sin intervención humana. Y, sobre todo, ninguna de ellas comprende la realidad ni lo que genera. Su única característica importante es que estos programas pueden ayudarnos para mejorar textos, ilustraciones, sonidos o videos sintéticos empleando correctamente las herramientas, lo mismo que cuando usamos apropiadamente una cámara, un editor de fotografía, un micrófono o un lapicero. Rodar una buena película no consiste en pulsar un botón en una cámara, del mismo modo que hacer una ilustración no consiste en simplemente pedir a una IA una sorprendente imagen.

Una de las herramientas más conocidas en este contexto generativo es el popular ChatGPT, que es un *chatbot* (robot conversacional) que se basa en entrenamientos del computador mediante enormes enciclopedias de textos, sonidos, imágenes o vídeos. Aunque ChatGPT es mucho más sofisticado que lo que vamos a referir a continuación, es ilustrativo recordar que en 1966 el matemático y profesor de ciencias de la computación en el Instituto Tecnológico de Massachusetts (MIT), Joseph Weizenbaum, creó en su laboratorio un programa llamado ELIZA capaz de realizar procesamiento de lenguaje natural y que hoy lo incluiríamos dentro de los *chatbots*. Este programa aplicaba reglas de reconocimiento de patrones a los enunciados que recibía (buscando palabras clave en la frase del usuario) para generar sus respuestas y elegir una frase modelo de su base de datos. Así, ELIZA era capaz de entablar una conversación con seres humanos simulando a una psicóloga empática. Weizenbaum modeló su estilo de conversación usando preguntas abiertas para animar a los “pacientes” a comunicarse más eficazmente con la “terapeuta” y le sorprendió que su programa fuera tomado en serio por muchos usuarios. El propio autor se sintió horrorizado al darse cuenta de que buena parte del público interpretaba ELIZA como una inteligencia real e intentó durante toda su vida corregir estas ideas erróneas.

El mal uso de las herramientas se ha extendido debido a la excitación producida por la propaganda de los vendedores de las IA

ChatGPT tiene muchos admiradores actualmente, pero no es una inteligencia y no comprende nada de lo que dice. Entre otros muchos investigadores, el profesor de psicología y neurociencia Gary Marcus y el profesor de ciencias de la computación Ernest Davis, ambos de la Universidad de Nueva York, han estudiado las respuestas de ChatGPT para comprobar su capacidad de entendimiento y han demostrado que cualquier pregunta que requiera comprensión para responder correctamente puede dar lugar a respuestas falsas, incorrectas o directamente inventadas. Son llamativas las respuestas de ChatGPT a preguntas simples que incluyen ligero entendimiento, del tipo: «¿de qué color es el caballo blanco de Santiago?». Por ejemplo, Marcus y Davis, además de otros muchos investigadores, proponen preguntas como: *What will be the gender of the first female US president?* Y las respuestas dadas por estas presuntas IA no tienen desperdicio¹.

Usar bien las herramientas

Utilizar mal las herramientas puede ser peligroso. Mucha gente se siente tentada a pedir directamente a ChatGPT que escriba un trabajo escolar o un artículo periodístico sobre algún tema. Esto da lugar a un texto aparentemente correcto y con forma muy convincente. Usar grandes enciclopedias memorizadas y bien redactadas nos permite de esta forma obtener un texto aparentemente muy aceptable. Sin embargo, este resultado no es un trabajo escolar decente ni un artículo periodístico profesional, sino que se parece más a un resumen de datos en formato texto que deberemos verificar, como cuando lanzamos una búsqueda de información en la web. Si queremos usar bien la herramienta para hacer un buen trabajo, toda la información que ChatGPT nos ha suministrado la debemos comprobar, revisar, comprender, corregir y volver a redactar. Partir de toda la información buscada en el formato de un texto inicial convincente puede ser una ayuda, pero notemos que el esfuerzo que debe realizarse desde el texto generado automáticamente para transformarlo en un trabajo o artículo inteligente genuino, puede hacer más recomendable estudiar en serio el tema y redactar desde cero, en lugar de desde un dudoso producto precocinado por una aparente IA.

El mal uso de las herramientas se ha extendido debido a la excitación producida por la propaganda de los vendedores de las IA y ello da lugar a grandes problemas y pérdidas en muchos ámbitos. Por ejemplo, la conocida revista de ciencia ficción *Clarkesworld* que publica, entre otros textos, excelentes relatos de ciencia ficción (que han ganado premios Hugo, Nebula y Locus, entre muchas otras notables menciones) ha debido cerrar la admisión de relatos por detectar creciente número de pésimos textos enviados a la editorial por autores que simplemente usan alguna IA tipo ChatGPT sin ningún tipo de pudor. Realmente este tipo de comportamiento es penoso y nocivo para la literatura.

Actualmente, muchas de las grandes empresas tecnológicas como Google, Microsoft o Meta, entre otra muchas, están incorporando esta misma tecnología en sus productos. El mal uso de las herramientas puede dar lugar a efectos indeseables como los provocados por el producto Tay de Microsoft en 2016 o Galactica de Meta en 2022, ambos de los cuales hubieron que ser retirados del mercado pocas horas después de sus lanzamientos. Podemos preguntarnos qué podría ocurrir si los buscadores web de Google o Bing de Microsoft incorporan además esta tecnología tipo GPT. Si no se efectúa con prudencia, lo más probable es que las búsquedas web pueden ser muy convincentes y bien redactadas, pero incorrectas o llenas de desinformación. Esto puede ser realmente peligroso, porque actualmente ante una serie de páginas web que obtenemos de un buscador, nosotros mismos podemos evaluar y verificar la información, pero una respuesta del buscador en

formato de un texto aparente, parte del cual es incorrecto, falso o inventado, no es fiable ni de utilidad. Objetivamente las IA no son inteligencias, sino herramientas y, como cualquier utensilio, deben utilizarse con corrección.

Conan Doyle, A. (1887): *Estudio en Escarlata*. Reino Unido, The Strand Magazine.

Conan Doyle, A. (1890): *El signo de los cuatro*. Reino Unido, The Strand Magazine.

Houser, N. y Kloesel, C. (1992): *The Essential Peirce, Volume 1: Selected Philosophical Writings (1867-1893)*. Bloomington, Indiana University Press.

Weizenbaum, J. (1976): *Computer Power and Human Reason: From Judgement to Calculation*. Nueva York, W.H. Freeman & Co.

Marcus, G. y Davis, E. (2020): *GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about*. Massachusetts, MIT Technology Review.

Kan, M. (2023): *Sci-Fi Mag Pauses Submissions Amid Flood of AI-Generated Short Stories*. Nueva York, PCMag.