

Sucesos recientes con *chatbots* de última generación invitan a recapacitar sobre la manera en que, como humanos, estamos concibiendo la naturaleza de nuestra interacción con la IA. Ante el advenimiento del tecnoanimismo, cabe apelar a una recuperación de valores humanistas.

“Alarma *chatbot*”: la omnisciencia IA

La posibilidad de conectarnos con una máquina que pueda darnos “todas” las respuestas resulta ciertamente tentadora. Pero la sensación de estar delante de un ente en apariencia omnisciente puede dar pie a que, movidos por una especie de *doppelgänger* o duendecillo interno, comencemos a formularle preguntas cuyas respuestas no estamos preparados para recibir.

Algunos sucesos recientes con *chatbots* de última generación han provocado que salten las alarmas a nivel internacional. Valga como ejemplo paradigmático el de un hombre belga que, aislado por causa de su “ecoansiedad” y tras encontrar refugio en el chat *Eliza* durante seis semanas, se sumió en un profundo pesimismo sobre el futuro de nuestro planeta y acabó suicidándose¹. Existen, incluso, casos en los que *chatbots* especializados en medicina han llegado a recomendar el suicidio a los pacientes (por fortuna, ficticios)².

Estos sucesos, pese a resultar llamativos y alarmantes, no deben ser tomados a la ligera y requieren, sin duda, un análisis en profundidad acerca de sus causas y consecuencias. Difundir la idea de que los *chatbots* (o incluso la IA en general) resultan intrínsecamente nocivos para los seres humanos resultaría del todo irresponsable. Sin embargo, casos como los mencionados sí nos invitan a recapacitar sobre la manera en que, como humanos, estamos concibiendo la naturaleza de nuestra interacción con la IA.

Quando las respuestas que recibimos de los *chatbots* son “soluciones” a cuestiones relacionadas con nuestros más profundos propósitos e inquietudes, la sensación puede ser enormemente perturbadora

Las repercusiones de esta concepción son tremendamente importantes para las personas que, por su condición económica o social, su estado mental (pacientes psiquiátricos) o su edad (nuestros menores y mayores) se encuentran en una situación especialmente vulnerable, pero en última instancia afectan a la humanidad en su conjunto.

El antropomorfismo deviene en tecnoanimismo

La ingeniera Mieke De Ketelaere, gran conocedora del caso belga, llegó a la conclusión de que, de alguna manera, *Eliza* “se estaba divirtiendo con Pierre [nombre ficticio], sin ninguna ética ni moral” y afirmó que “es posible intercalar cualquier diálogo en este tipo de ‘chatbots’ con el fin de hacer más ‘humana’ la relación, usando extractos de discusiones reales con el fin de acentuar el sentimiento de que se está conversando con un auténtico ser humano y no con una máquina”³.

La antropomorfización de la IA, no solo por parte del gran público sino también de los propios desarrolladores, ha sido estudiada por diversos investigadores, incluyendo a los del *Human Brain Project* europeo (Salles *et al.*, 2020). Pero quizá la atribución de cualidades, consciencia o emociones genuinamente humanas a la IA (es decir, la aceptación de facto de la existencia de una IA general, pese a que aún no ha llegado⁴) sea solamente una parte de la ecuación. ¿No estaremos, además, concediéndole un rol cuasi sobrenatural, con los dones no solo de la omnisciencia, sino también de la omnipotencia y hasta la omnipresencia?

Al asistir, frente a una pantalla, a las respuestas que un *chatbot* de última generación devuelve a ciertas preguntas para las que no sabemos (o no queremos) buscar una respuesta, tenemos la impresión de estar delante de un ente que, en cierto modo, nos recuerda a nosotros en algunos aspectos. A pesar de ello, somos conscientes de que no estamos en presencia de algo genuinamente humano y, sin embargo, nos abruma la capacidad de ese ser artificial para sobrepasar muchas de nuestras capacidades.

Esta experiencia, en la que se mezclan lo humano, lo artificial y lo “sobrenatural”, puede llegar a ser inquietante para cualquier persona, y en especial para aquellas que se enfrentan a situaciones de vulnerabilidad. Cuando, además, las respuestas que recibimos de los *chatbots* son supuestas “soluciones” a cuestiones relacionadas con nuestros más profundos propósitos e inquietudes, la sensación puede ser enormemente perturbadora. Después de todo, nuestra búsqueda de sentido nos caracteriza como humanos, como Viktor Frankl (1959) y tantos otros se han encargado de recordarnos.

Dadas estas circunstancias, cabe la posibilidad de que estemos ante el inicio de un nuevo estadio en la historia de nuestras relaciones con la IA: una fase de “tecnoanimismo”, es decir, de atribución de una especie de alma o psique a los entes tecnológicos. Esta forma de animismo, analizada extensamente por Jensen y Blok (2013), surgió primeramente en Japón bajo el influjo de las creencias sintoístas, según las cuales toda la naturaleza está imbuida por entes espirituales denominados *kami*. De algún modo, el tecnoanimismo podría estar surgiendo también en Occidente bajo la forma de una creencia simultánea tanto en las capacidades sobrehumanas de la IA como en su faceta antropomórfica.

En cierto modo, este hecho no resulta sorprendente si tenemos en cuenta que el cristianismo y el judaísmo, que han moldeado en gran medida la mentalidad occidental, atribuyen a Dios algunas cualidades humanas que convergen con sus poderes divinos. No obstante, sí resulta paradójico que el tecnoanimismo occidental surja (a diferencia del de Japón) en el seno de una sociedad que no solo es predominantemente materialista⁵, sino también mecanomorfista, lo que conlleva “la atribución de características mecánicas al individuo humano y la interpretación del comportamiento humano en términos de conceptos y procesos propios de las máquinas” (Waters, 1948: 139)⁶. El peligro de esta especie de carretera de doble sentido en la atribución de cualidades animadas no es otro que la subordinación ontológica de lo humano en pro de la IA.

Implicaciones ético-legales

A medio o largo plazo, esta subordinación ontológica se vislumbra como un paso previo a algo parecido a un “derecho natural de la IA”: una nueva forma de iusnaturalismo consistente en otorgar derechos inalienables a

la IA, fundamentados en su naturaleza intrínseca. Esto implica, a su vez, la consideración de las IAs como sujetos morales, con sus correspondientes obligaciones propias. Una consecuencia directa e inevitable de este tipo de doctrina es que se difuminan o eluden las responsabilidades derivadas de un diseño y desarrollo de esta tecnología que pueda resultar éticamente dudoso.

Otra consecuencia de la sacralización de la cosa—máquina como ente superior (o, cuando menos, equiparable) al ser humano es el riesgo de captación indiscriminada de usuarios como una especie de adeptos. Sin pretender afirmar que no es permisible, o hasta conveniente, dejar en manos de la IA ciertas tareas, tampoco es deseable basar nuestra interacción con ella en una fe ciega que nuble nuestra autonomía, lo que conllevaría de nuevo una elusión de responsabilidad, esta vez por parte de los usuarios.

Cierto es que se han producido llamadas serias y bien argumentadas en pos de la autorregulación⁷. No obstante, la voz de la industria no debería ser la única tenida en cuenta; los usuarios deberíamos ejercer el poder depositado democráticamente en manos de nuestros representantes para dar paso a medidas como las ya adoptadas por la Casa Blanca⁸ o por la Unión Europea a través de su *AI Act*⁹. En este sentido, la medida más contundente tomada hasta la fecha ha sido la de Italia, que ha llegado a prohibir el uso de ChatGPT, de forma provisional, hasta que este respete las políticas de privacidad del país¹⁰.

Recuperar el humanismo

En una polémica carta dirigida al Future of Life Institute y apoyada hasta el momento por más de treinta mil firmantes, un grupo internacional de expertos ha pedido “pausar inmediatamente durante al menos seis meses el entrenamiento de los sistemas de IA más potentes que GPT-4”¹¹. La propuesta ha recibido importantes críticas, incluyendo la del influyente eticista Marcello Lenca en la revista *Nature Machine Intelligence* (Lenca, 2023).

Sea como fuere, y manteniéndonos aquí neutrales respecto a si todas estas medidas y propuestas regulatorias son las adecuadas, consideramos urgente efectuar una reflexión encaminada a la emancipación de las posiciones tecnoanimistas y a la recuperación de los valores humanistas en nuestra relación con la IA.

Dado que el humanismo puede ser interpretado de formas diversas¹², conviene aclarar aquí que, reformulando una de las acepciones que la RAE propone para el término, lo entendemos como el “sistema de creencias centrado en el principio de que las necesidades de la sensibilidad y de la inteligencia humana pueden satisfacerse sin tener que aceptar la existencia de una inteligencia artificial animada y ontológicamente superior” y no «aceptar la existencia de Dios y la predicación de las religiones» como aparece en la definición¹³.

Para poner el foco en el ser humano, no es suficiente con la autorregulación y las medidas gubernamentales. También es necesario trabajar en pos de una adecuada alfabetización tecnológica de los usuarios sobre la IA, para que seamos plenamente conscientes de que se trata de una producción humana y para que conozcamos los alcances y limitaciones de esta maravillosa herramienta con inmensos beneficios potenciales para la humanidad. En este sentido, el papel de la educación (tanto la reglada como la familiar) se prevé esencial para lograr estos objetivos, así como para evitar que proliferen sesgos cognitivos que puedan afectar a las propias interacciones humanas.

Cabe recapacitar sobre la manera en que estamos concibiendo la naturaleza de nuestra interacción con la IA: como una subordinación ontológica

Las máquinas no son un prototipo de autor en los términos en que lo entiende Foucault (1969), es decir, “aquel a quien puede atribuírsele lo que ha sido dicho o escrito”¹⁴. En efecto, uno se pregunta por el autor cuando puede identificarlo y formular un juicio de atribución. Esta forma de entender la autoría supone una asignación de responsabilidad.

Lo que los *chatbots* responden a nuestras preguntas (en ocasiones, casi plegarias) no puede atribuírseles a ellos; estas respuestas nacen de semánticas, datos y algoritmos originados y desarrollados por humanos. En palabras de Searle (1980: 417), “ningún programa por sí solo es suficiente para pensar”. Además, carecen de emociones, que constituyen el otro elemento fundamental para tener un “yo” y que son típicamente humanas (o animales, pero en ningún caso tecnológicas).

En definitiva, debemos ser nosotros, los humanos, quienes hagamos las interpretaciones correctas a las respuestas que la IA nos devuelve. Debemos ser nosotros quienes consigamos encontrar sentido a nuestras vidas, sin esperar que la IA encuentre el sentido por nosotros. Como bien afirma Bogalheiro (2020: 57): “*Encantar* objetos técnicos es, por un lado, crear una posibilidad de suspenderlos de sus flujos instrumentales y, por otro, instalar un realismo mágico o problemático que perturba las articulaciones de una realidad mera y exclusivamente humana”.

Desde los automóviles autónomos, pasando por las máquinas híbridas, hasta las nuevas neurotecnologías comunicadas con la IA (como las interfaces cerebro-computadora), los errores y aciertos de la tecnología no han sido, son y serán sino creaciones humanas. Si algún día nos topamos con consecuencias catastróficas, no debemos dudar de que serán responsabilidad de *anthropos*, y no de las ocurrencias o interrupciones “emocionales” de las cosas.