ANALES
RANF
www.analesranf.com

EPISTEMOLOGÍA Y ONTOLOGÍA EN CIENCIA: EL RETO DE LA INTELIGENCIA ARTIFICIAL

EPISTEMOLOGY AND ONTOLOGY IN SCIENCE: THE CHALLENGE OF ARTIFICIAL INTELLIGENCE

Santiago Cuéllar Rodríguez
Académico correspondiente de la Real Academia Nacional de Farmacia. ORCID: https://orcid.org/0000-0002-8099-9226

corresponding author: santiago.cuellar.rodriguez@gmail.com

ARTÍCULO DE REVISIÓN

RESUMEN

La brecha entre predictibilidad y comprensibilidad amenaza todo el proyecto científico porque los modelos matemáticos de los procesos, alimentados por enormes cantidades de datos de origen muy diverso, proporcionan resultados excepcionalmente precisos pero, al mismo tiempo, ocultan la explicación de los procesos. El conocimiento de "qué sabemos" de la ontología es tan relevante en ciencia como el de "cómo sabemos" y el de "cuánto sabemos" de la epistemología. La inteligencia artificial (IA) implica la comprensión científica de los mecanismos que subyacen al pensamiento y la conducta inteligente, así como su encarnación en máquinas capacitadas por sus creadores de razonar en un sentido convencional. Su formulación "débil" se refiere al empleo de programas informáticos complejos, diseñados con el fin de complementar o auxiliar el razonamiento humano para resolver o completar complejos problemas de cálculo, de mantenimiento de sistemas, de reconocimiento de todo tipo de imágenes, de diseño, de análisis de patrones de datos, etc., muchos de los cuales serían prácticamente inabordables mediante procedimientos convencionales; pero todo ello sin incluir capacidades sentientes o éticas humanas, que sí serían objeto de una – por ahora – inexistente IA "fuerte", aquella que igualaría o incluso excedería la inteligencia sentiente humana. La vulgarización de la IA "generativa", desarrollada para crear contenido – texto, imágenes, música o vídeos, entre otras muchas áreas – a partir de información previa, está contribuyendo a consolidar popularmente la idea errónea de que la actual IA excede el razonamiento a nivel humano y exacerba el riesgo de transmisión de información falsa y estereotipos negativos a las personas. Los modelos de lenguaje de la inteligencia artificial no funcionan emulando un cerebro biológico sino que se fundamentan en la búsqueda de patrones lógicos a partir de grandes bases de datos procedentes de fuentes diversas, que no siempre están actualizadas ni depuradas de falsedades, de errores ni de sesgos conceptuales o factuales, tanto involuntarios como interesados. Y la IA empleada en ciencia no es ajena a estas limitaciones y sesgos. Una cuestión particularmente sensible es la posibilidad de utilizar la IA generativa para redactar o incluso inventarse artículos científicos que llegan a pasar desapercibidos por los revisores por pares de las revistas científicas más prestigiosas del mundo, apuntando a un problema más aún profundo: los revisores por pares de las revistas científicas a menudo no tienen tiempo para revisar los manuscritos a fondo en busca de señales de alerta y, en muchos casos, además carecen de recursos informáticos adecuados y formación especializada.

Palabras Clave:

ontología
epistemología
método científico
inteligencia artificial
predictibilidad
comprensibilidad

Keywords:

ontology
epistemology
scientific method
artificial intelligence
predictability
understandability

ABSTRACT

The gap between predictability and comprehensibility threatens the entire scientific project because mathematical models of processes, fed by enormous amounts of data of very diverse origin, provide exceptionally precise results but, at the same time, hide the explanation of the processes. The knowledge of "what we know" of ontology is as relevant in science as that of "how we know" and "how much we know" of epistemology. Artificial intelligence (AI) involves the scientific understanding of the mechanisms underlying intelligent thought and behavior, as well as their embodiment in machines trained by their creators to reason in a conventional sense. Its "weak" formulation refers to the use of complex computer programs, designed with the purpose of complementing or assisting human reasoning to solve or complete complex problems of calculation, system maintenance, recognition of all types of images, design, analysis of data patterns, etc., many of which would be practically unapproachable using conventional procedures; but all this without including human sentient or ethical capabilities, which would be the subject of a – at the moment – non-existent "strong" AI, that would equal or even exceed human sentient intelligence. The popularization of "generative" AI, developed to create content – text, images, music or videos, among many other areas – from previous information, is helping to popularly consolidate the erroneous idea that current AI exceeds reasoning human level and exacerbates the risk of transmitting false information and negative stereotypes to people. The language models of artificial intelligence do not work by emulating a biological brain but are based on the search for logical patterns from large databases from diverse sources, which are not always updated or purged of falsehoods, errors or errors. conceptual or factual biases, both involuntary and self-serving. And the AI used in science is no stranger to these limitations and biases. A particularly sensitive issue is the possibility of using generative AI to write or even invent scientific articles that go unnoticed by the peer reviewers of the most prestigious scientific journals in the world, pointing to an even deeper problem: peer reviewers. Reviewers often do not have the time to review manuscripts thoroughly for red flags and, in many cases, they also lack adequate computing resources and specialized training.

1. INTRODUCCIÓN

Llamamos epistemología a la rama de la filosofía dedicada al estudio de la naturaleza, las fuentes y los límites del conocimiento: qué es, cómo se produce, qué podemos conocer, etc. Responde a la necesidad de comprender racionalmente en qué se basan y cuál es el grado de confianza en nuestras convicciones profundas y cuáles son las limitaciones de tal confianza, impuestas por la evidencia empírica y por la razón; en definitiva, por qué y en qué grado estamos convencidos lo que estamos convencidos.

En ciencia los valores epistémicos son las propiedades de una hipótesis, teoría o convicción que son indicativas de su veracidad y que están directamente vinculados con la búsqueda del conocimiento; por ello, tienen que ver con aspectos tan relevantes como la exactitud, la consistencia, el alcance, la simplicidad, la objetividad, la reducción de errores, la eficacia, la robustez, el poder predictivo, la novedad, la aplicabilidad, la uniformidad ontológica, el poder explicativo y la coherencia externa, entre otros.

Hace algún tiempo revisamos (1) algunos aspectos de la epistemología, especialmente con relación a qué valor tienen, cómo se generan y cómo se validan los conocimientos proporcionados por la medición en la ciencia. Decíamos entonces que la brecha entre predictibilidad y comprensibilidad amenaza con volar por los aires todo el proyecto científico debido a que parece que hemos llegado a un límite en el que la comprensión y la predicción se están desalineando (2), entre otros motivos porque el crecimiento exponencial del "big data" y, muy particularmente, de la inteligencia artificial (IA) está acrecentando esa brecha entre predictibilidad y comprensibilidad. Los modelos matemáticos, alimentados por monstruosas cantidades de datos de origen muy diverso, proporcionan predicciones cuasi-milagrosas pero, al mismo tiempo, nos ocultan la explicación hasta el punto de llevarnos a pensar que la razón casi llega a ser innecesaria, siempre que las predicciones merezcan la pena.

En contraposición a la epistemología, Robert Proctor y Londa Schiebinger (3) acuñaron el término "agnotología", para designar el procedimiento de investigar la creación de ignorancia y de sus numerosos y diversos mecanismos de producción, tales como el secreto, la confusión, la pereza en la búsqueda, el desprecio de las fuentes fiables y el ominoso silencio cómplice, entre otros. Proctor y Schiebinger nos avisan de que la ignorancia está, a veces, solapada con la ciencia y con la razón; "la ignorancia se esconde en las sombras de la filosofía".

Hasta hace unas pocas décadas, se había mantenido un delicado equilibrio entre la ontología, es decir, la comprensión y el conocimiento profundo de la naturaleza de las cosas y de los fenómenos del mundo, y su medida y predicción, la epistemología, que se ocupa de los procesos mediante el que adquirimos y validamos tal conocimiento. Ese equilibrio buscaba optimizar la comprensión, lo cual nos permitía apreciar nuevas características y formular mejores leyes fundamentales de la realidad que, a su vez, facilitaban hacer nuevas predicciones para confirmarlas o refutarlas, completando un riguroso procedimiento científico.

Sin embargo, en la actualidad, tal equilibrio se ha ido desplazando progresivamente hacia la parte experimental y, de forma muy particular, hacia la computación de ingentes cantidades de datos, muchas veces no adecuadamente depurados ni por la credibilidad de su origen ni por la homogeneidad de sus significados. Es cierto que, de otra manera, seguiríamos atascados en problemas de muy difícil solución. La mecánica cuántica es un ejemplo paradigmático de ello y cómo hace un siglo vino para justificar y predecir resultados que de otra manera "nos mantendrían atascados en el barro". Pero, a cambio de unos resultados extraordinariamente exactos, la mecánica cuántica nos impuso como tributo no poder intuir el mecanismo que los produce, como reconocían algunos de sus científicos más significados, tales como que "la mecánica cuántica es fundamentalmente incomprensible" (Niels Bohr), o que "si crees que entiendes la mecánica cuántica, es que no entiendes la mecánica cuántica" (Richard Feynman), hasta llegar a la abnegada aceptación de John von Neumann de que "no entiendes la mecánica cuántica, simplemente te acostumbras" o a la irritada exclamación de David Mermin de "¡Cállate y calcula!", para acatar sin rechistar una interpretación instrumentalista – un formalismo matemático puramente simbólico – que simplemente funcione: no importa que la teoría no nos informe sobre la realidad, en tanto que los datos calculados encajen adecuadamente con los resultados de los experimentos.

Pese a todo, el conocimiento de "qué sabemos" de la ontología es tan relevante en ciencia como el de "cómo sabemos" y el de "cuánto sabemos" de la epistemología. Incluso en áreas de la investigación científica donde resalta el valor del cálculo numérico, como la estadística, su gran reto epistémico consiste en tratar de justificar el valor de unos procedimientos que, a partir de datos reales, permiten alcanzar veredictos predictivos o confirmatorios sobre la validez de una hipótesis, porque cuando la aritmética – o, en general, la matemática – se establece como una cárcel epistémica de la realidad pierde por completo todo su sentido, como nos ayudó a comprender Kurt Gödel con sus teoremas de indecibilidad o incompletitud.

En el estudio de cualquier fenómeno, la intervención del observador – el que realiza la medición – afecta al fenómeno estudiado; este reiterado "mantra" científico nos viene a recordar que la simple observación – y aún más la medición – siempre

produce algún grado de perturbación que limita o incluso llega a "congelar" la percepción de la realidad, haciéndola incompleta. Aunque esta visión es relativamente reciente en el pensamiento occidental, no lo es tanto en el oriental; en concreto, en la epistemología clásica china la autoconciencia – la capacidad de conectar con los propios sentimientos, pensamientos y acciones – de cada persona se basa en la comprensión global – holística – del mundo, entendida como una relación interactiva entre cada ser humano y la naturaleza, donde cada ser está incrustado y entretejido con estructuras cósmicas. En definitiva, nuestra comprensión de la existencia está ligada a la forma de cómo experimentamos la existencia.

En medicina y, en general, en la ciencia existe una fuerte interrelación entre evidencia (los hechos observados y contrastados), los valores epistémicos (propiedades de una hipótesis o teoría que son indicativas de su veracidad y que están directamente vinculados con la búsqueda del conocimiento), los valores no epistémicos (aparentemente desconectados de la búsqueda del conocimiento, pero necesarios para determinar lo que se considera como una prueba suficiente: aspectos morales, políticos, sociales, personales, económicos, etc.) y los sesgos cognitivos (procesos mentales que se desvían sistemáticamente de las normas reconocidas de la lógica y la racionalidad y, al hacerlo, afectan a nuestro juicio y a la correspondiente toma de decisiones). Tal interrelación debe ser equilibrada porque si no puede conducir a lo que Cristina Amoretti (4) denomina "una pandemia de tonterías" en el ámbito de la salud pública, como las disparatadas y contradictorias medidas de salud pública que fueron adoptadas por los gobiernos de la inmensa mayoría de los países durante la pandemia de COVID-19. Por ello y con el fin de restaurar y promover la confianza pública en la medicina y en toda actividad científica, considera Amoretti que tal interacción entre evidencia, valores (epistémicos y no epistémicos) y sesgos cognitivos debe hacerse explícita y discutirse públicamente. En ciencia, una prueba es definida como una construcción metódica que permite confirmar o descartar significativamente una hipótesis, y que puede ser revisada racionalmente, reproducida y verificada por cualquier persona ajena al estudio original que disponga de los conocimientos y los medios adecuados para ello. Para Andrew Granville (5) el mejor sistema de verificación que tenemos en matemáticas es que muchos expertos analicen la prueba desde diferentes perspectivas y que, en todas ellas, encaje bien. Esto no demuestra la veracidad irrefutable de la prueba, pero confirma a la comunidad científica que al menos es correcta o, lo que es lo mismo, que no encontramos ningún elemento discordante que nos impida seguir avanzando por ese camino; de hecho, las pruebas son aceptadas según estándares científicos comunitarios y es precisamente la comunidad científica – en su amplia diversidad y libertad – la que proporciona confianza. Sin embargo, en demasiadas ocasiones se olvida – con la complicidad de no pocos editores científicos – que la crítica racional libre y diversa es una condición fundamental para considerar como científico a un estudio. Miranda Fricker (6) ha mostrado que los estereotipos, las relaciones sociales jerárquicas y los roles de género tradicionales afectan la forma en que se adquiere y difunde el conocimiento, pero también a cómo los demás perciben a sus creadores. Fricker acuñó el término "injusticia epistémica" para referirse a aquella que implica que una persona o un colectivo reciba poco o ningún crédito debido a su identificación con un estatus social que es consecuencia casi siempre de posiciones crónicamente desfavorecidas, pese a que pueden tener una visión privilegiada en numerosas materias debido – precisamente – a su condición de marginación.

El desarrollo científico no es un modelo uniforme de evolución del conocimiento y, desde luego, no se debe exclusivamente al seguimiento de procesos racionales lógicos ni de rigurosos y precisos protocolos de metodología práctica. Procedimientos tan comunes como poco estudiados como la intuición nos retan a comprender por qué son fundamentales en el desarrollo de la ciencia, como nos recordaba Henri Poincaré, para quien la intuición proporciona "una base epistemológica a priori para las matemáticas" y supone la consciencia de poseer una capacidad innata para construir un continuo físico y matemático, una capacidad sin la que la experiencia humana se reduciría a un conjunto incoherente de meras sensaciones aisladas. Podemos pensar que, aplicado a otros ámbitos del conocimiento y de la comprensión, el concepto de intuición podría traducirse como la experiencia que proporciona una visión de rayos X que, aunque a veces es errónea, casi siempre permite alcanzar una conclusión mucho antes de tener pruebas objetivas que la avalen; es el ojo clínico del buen médico o el olfato del periodista independiente, que les permite ir más allá de la apariencia de lo evidente. Una especie de "túnel cuántico" que atraviesa las cordilleras más altas del conocimiento sin necesidad de llegar a sus cumbres escalando a través de escarpadas pendientes.

Las matemáticas, decía Poincaré hace un siglo, "requieren intuición no solo en el contexto del descubrimiento sino también en el contexto de la justificación, especialmente en aritmética y lógica". Por eso la matemática Eugenia Cheng (7) afirma hoy que "en la investigación matemática, no solo sigues pasos lógicos, si lo haces nunca llegarás a ningún lugar interesante; tienes que usar tu instinto y sentir tu camino a través de algo primero, y respaldarlo con lógica después".

No solo en ciencia sino en el ámbito de la vida cotidiana, la mayoría de nosotros – probablemente sin darnos cuenta – frecuentemente compartimos falsedades triviales. Solo en algunos

casos nos tomamos la molestia de verificar y reproducir los hechos o fundamentos de nuestras datos y argumentaciones empleadas. Pero, como indica Richard Reeves (8), el problema no consiste simplemente en poder discernir lo verdadero de lo falso, sino en quién quiere decir la verdad; la cuestión no es tanto dónde está la verdad como quién es sincero. Como el propio Reeves reflexiona, el carácter de la verdad es empírico, mientras que el de la veracidad es ético; la verdad es el producto final pero la veracidad es un elemento esencial en su producción y difusión. La temida crisis epistémica de la racionalidad – la que afecta al saber construido con una metodología rigurosa – es, por tanto, una crisis ética que requiere soluciones éticas.

2. ONTOLOGÍA Y EPISTEMOLOGÍA EN INTELIGENCIA ARTIFICIAL (IA)

La inteligencia artificial (IA) es definida por la Association for the Advancement of Artificial Intelligence (AAAI) como "la comprensión científica de los mecanismos que subyacen al pensamiento y la conducta inteligente, y su encarnación en máquinas; es decir, la capacidad para razonar por parte de un agente que no está vivo en un sentido convencional, al cual le ha sido conferida dicha capacidad por seres humanos. Además de razonar, estos dispositivos son – potencialmente – capaces de desarrollar otras actividades tradicionalmente consideradas como humanas".

Como tal, la definición resulta un tanto ambigua, susceptible de sostener, al menos, dos formulaciones diferentes. La que podríamos llamar formulación "débil" se refiere al empleo de programas informáticos complejos, diseñados con el fin de complementar o auxiliar el razonamiento humano en la resolución de problemas específicos, aunque sin incluir capacidades sentientes o éticas humanas. Esta forma "amable" de la inteligencia artificial nos ayuda extraordinariamente a resolver o completar complejos problemas de cálculo, de mantenimiento de sistemas, de reconocimiento de todo tipo de imágenes, de diseño, de análisis de patrones de datos, etc., que, de otra manera, muchos de los cuales serían prácticamente inabordables por mediante los procedimientos convencionales (9). Por su parte, la inteligencia artificial "fuerte" es aquella que iguala o excede la inteligencia humana promedio, pudiendo – idealmente – realizar con éxito cualquier tarea intelectual de un ser humano.

De momento, solo podemos considerar como una realidad a la forma "débil" de inteligencia artificial, aunque las capacidades sobrevaloradas, poco realistas y exageradas impregnan la forma en que se presentan públicamente los modelos de IA "generativa" (desarrollada para crear contenido – texto, imágenes, música o vídeos, entre otras muchas áreas – a partir de información previa), lo que contribuye a la idea errónea de que estos modelos exceden el razonamiento a nivel humano y exacerba el riesgo de transmisión de información falsa y estereotipos negativos a las personas; de hecho, la fabricación y el sesgo en los modelos de inteligencia artificial (IA) generativa pueden ocurrir como parte del uso regular del sistema, incluso en ausencia de fuerzas malévolas que busquen impulsar el sesgo o la desinformación (10).

Un ejemplo paradigmático de la IA generativa son los bot de charla o bot conversacionales (en inglés, chatbot), aplicaciones informáticas que simulan mantener una conversación con un ser humano, proveyendo respuestas automáticas generadas a partir de un amplio conjunto de datos y una arquitectura informática definida por sus programadores, utilizando lo que se denomina como "razonamiento basado en casos" (case base reasoning, CBR). Aunque los resultados son a veces muy espectaculares, conviene no olvidar que, en cualquier modelo de inteligencia artificial, los sesgos que existen en los conjuntos de datos que son empleados para entrenar y mantener el sistema se transmiten inevitablemente al comportamiento del chatbot. Por eso, con el fin de distinguir entre los textos generados por sistemas de inteligencia artificial y los creados por seres humanos, algunos desarrolladores informáticos están empleando lo que se conoce como "análisis de la perplejidad", consistente en determinar la aleatoriedad, el grado de desorden que hay en un texto; de tal manera, una perplejidad alta indicaría mayor probabilidad de que el texto haya sido generado por una persona y no por una máquina. Pero hay otro método de reconocimiento que nunca falla: la inteligencia artificial no entiende los "porque sí", ni se contradice a sí misma, como frecuentemente hacemos los seres humanos.

Los modelos de lenguaje de la inteligencia artificial no funcionan emulando un cerebro biológico, sino que se fundamentan en la búsqueda de patrones lógicos – definidos por programadores humanos – dentro de descomunales bases de datos suministrados a partir de diversas fuentes, no siempre actualizadas ni depuradas de falsedades, errores y sesgos conceptuales o factuales, tanto involuntarios como interesados. Por tanto, es previsible que algunas de sus respuestas estén contaminadas por esos "pecados", si bien los filtros y algoritmos empleados son cada vez más cuidadosos a este respecto.

Más allá de su eficacia como potencial generador de conocimiento, existe una preocupación: ¿Qué argumentos nos convencerían de que una inteligencia artificial ha llegado a adquirir algún grado de "sensibilidad humana"? Mucho me temo que tales argumentos, en caso de existir, ya estarían implícitos – incrustados, de forma inconsciente o no – en el tipo y la forma de los datos que nosotros mismos aportamos a la inteligencia artificial para que "nos

convenza de que es sentiente". En realidad, tales argumentos solo emulan las formas en que los humanos expresamos nuestra afectividad y el resto de nuestros sentimientos; las máquinas solo recogen lo que los seres humanos sembramos, lo depuran, lo ordenan en una secuencia que los humanos consideramos lógica y, en realidad, solo acabamos leyendo lo que nosotros mismos hemos escrito. Es un ejemplo perfecto de argumento circular.

Pero, quizá, el principal problema no consista tanto en que las máquinas puedan acercarse al pensamiento sentiente humano, sino que los seres humanos nos estamos acostumbrando a usar de forma "natural" lo que es un lenguaje artificial que nosotros mismos hemos desarrollado para manejar las máquinas y ahora tratamos – irracionalmente – de encontrarle a las máquinas el mismo sentido que nuestra inteligencia "natural" les dio originalmente. A pesar de lo que algunos de sus defensores peor informados afirman, la inteligencia artificial es incapaz – por el momento – de comprender el verdadero sentido de un texto o de un discurso; tan solo, en el mejor de los casos, es capaz de identificar sus componentes y traducirlos linealmente a su propio idioma mecánico y, eventualmente, crear un texto o discurso coherente a partir de ellos; es decir, es capaz de leer y escribir como un ser humano… pero sin entender lo que lee o dice. En realidad, los humanos aprendemos el sentido real del lenguaje a partir de las interacciones y la comunicación con los demás humanos, no deletreando textos. La ambigüedad natural del lenguaje ordinario, la extremada dependencia contextual de su significado concreto y, especialmente, la necesidad de contrastarlo con gran cantidad de conocimientos de carácter cotidiano, hace incompetente a la inteligencia artificial – por el momento, insisto – para un auténtico diálogo humano, algo que cualquier niño de cuatro años es capaz de mantener sin dificultad.

Otro de los problemas que pueden surgir de la inteligencia artificial no es que haga algo extraño o ajeno a las instrucciones de su programador, sino que haga lo que éste le dijo que tenía que hacer, pero sin que el programador fuese plenamente consciente del alcance y consecuencias de sus propias instrucciones. Las máquinas no son buenas o malas, sino que al cumplir las instrucciones, calculan todas las posibilidades ejecutivas, incluyendo aquellas que no llegó a prever su programador; de hecho, las máquinas siempre acaban yendo más allá de los deseos de sus torpes o irresponsables manipuladores. Si deseamos que nuestro coche vaya muy rápido para disfrutar de la agradable sensación de vértigo y de poder que da la velocidad, llegará un momento en que no seamos capaces de evitar las consecuencias adversas de ello, incluso aunque nuestro coche nos avise con antelación. Todo ello porque dedicamos mucho más esfuerzo a idealizar nuestros deseos que a considerar sus efectos y consecuencias reales.

Estamos acostumbrándonos a que la inteligencia artificial esté presente en todas las formas y aplicaciones del conocimiento, quizá sin considerar que, en definitiva, solo son programas informáticos y dispositivos regidos por lógicas de diversos órdenes, con capacidad para reciclar continuamente sus resultados, optimizándolos y generando nuevas asociaciones de datos e hipótesis no contempladas originalmente. No cuestiono las notables posibilidades que proporciona esta tecnología; sin embargo, me inquieta que estemos dejando mansamente el control en manos de máquinas que, en definitiva, están diseñadas por personas con manías, ideas, prejuicios, fobias y sentido existencial, que están financiadas – y controladas estrechamente – por corporaciones privadas o públicas con intereses no siempre confesables; me preocupa que depongamos a nuestra inteligencia natural – nuestro pensamiento sentiente – en beneficio de la no siempre predecible inteligencia artificial y de sus dueños reales, que no son sus programadores.

Una cuestión particularmente sensible es la posibilidad de utilizar la IA generativa para redactar o incluso inventarse artículos científicos que llegan a pasar desapercibidos por los revisores por pares de las revistas científicas más prestigiosas del mundo. En 2021, Guillaume Cabanac, un científico informático de la Universidad de Toulouse (Francia) publicó los resultados de un trabajo de investigación documental en el que se analizó la presencia de algunas frases extrañas – aparentemente incongruentes con el texto científico – en miles de artículos académicos, que él denominó como "frases torturadas" (11).

La búsqueda de Cabanac de artículos con frases absurdas o "torturadas" ya había comenzado en 2015, cuando comenzó a colaborar con Cyril Labbé, un científico informático de la Universidad de Grenoble Alpes en Francia. Labbé había desarrollado un programa para detectar galimatías en artículos informáticos generados automáticamente mediante SCIgen, un software creado inicialmente como una broma. El trabajo de Labbé llevó a las revistas a retirar más de 120 manuscritos. Hasta 2021, Cabanac y sus colegas, junto con voluntarios de la comunidad PubPeer, han identificado cerca de 400 frases torturadas en más de 2.000 artículos, incluidos los de revistas de editoriales conocidas como Elsevier y Springer Nature. La Fundación PubPeer es una corporación registrada en California (Estados Unidos) con estatus de organización sin fines de lucro y cuyo objetivo general es mejorar la calidad de la investigación científica permitiendo enfoques innovadores para la interacción comunitaria, como un servicio dirigido en beneficio de sus lectores y comentaristas, quienes crean su contenido (12).

La realidad es que cada vez son más los manuscritos científicos que no revelan la asistencia de la IA y que están pasando desapercibidos para los revisores por pares de las grandes editoras

científicas del mundo. Aunque la presencia de tales artículos que están escritos total o parcialmente de forma fraudulenta mediante software de computadora no son nada nuevo, sin embargo, hasta hace poco tiempo eran más fácilmente detectables ya que solían mostrar ciertos rastros sutiles, tales como patrones específicos de lenguaje o las ya mencionadas "frases torturadas". Pero actualmente si los "detectives" que investigan este fraude eliminan las frases repetitivas de ChatGPT, el texto del chatbot se vuelve más fluido y sofisticado, haciendo "casi imposible" de detectar el fraude científico y editorial (13).

El uso no revelado de ChatGPT y otras herramientas de IA no solo se ha identificado en artículos de revistas científicas, sino también en conferencias revisados por pares y en preimpresiones (manuscritos que no han pasado por revisión por pares). Algunos de los autores denunciados por un uso no declarado de chatbot se disculparon alegando que lo habían utilizado "para ayudar a crear el trabajo".

El problema de los artículos no divulgados producidos por IA en revistas apunta a un problema más aún profundo: los revisores por pares de las revistas científicas a menudo no tienen tiempo para revisar los manuscritos a fondo en busca de señales de alerta. En este sentido, Rune Stensvold, microbiólogo del Instituto Estatal del Suero en Copenhague, se encontró con el problema de las referencias falsas inventadas por la propia inteligencia artificial cuando un estudiante le pidió una copia de un artículo del que aparentemente Stensvold había sido coautor con uno de sus colegas en 2006. El artículo simplemente no existía. El estudiante le había pedido a un chatbot de IA que le sugiriera artículos sobre Blastocystis, un género de parásito intestinal, y el chatbot había improvisado una referencia con el nombre de Stensvold. "Parecía tan real", dice. Quizá, por ello, sería conveniente que la revisión de artículos científicos comenzase comprobando la sección de referencias".

Finalmente, atendiendo al posible carácter sentiente y ético de la inteligencia artificial, es preciso considerar que una cosa es conocer y saber manejar con efectividad la mecánica de un juego – con reglas establecidas que ordenan la persecución de un objetivo previamente definido – y otra, muy diferente, es la satisfacción que produce jugar e incluso ganar, o haber aprendido una nueva jugada o estrategia; es muy improbable que una máquina pretendidamente inteligente tenga el impulso de inventar un juego para disfrutar retando a su propia inteligencia, ni que se aburra jugando con él. Asimismo, la voluntad libre para hacer algo, la independencia de los objetivos que le fijen es otro de los aspectos de los que carece la inteligencia artificial, porque ésta arrastra la doble condena de servir para un fin que ella misma no ha elegido y de no experimentar la alegría de conseguir sus metas ni la frustración de no haberlo hecho.

3. CONCLUSIONES

La brecha entre predictibilidad y comprensibilidad amenaza todo el proyecto científico debido a que parece que estamos llegando a un límite en el que la comprensión y la predicción se están desacoplando. Los modelos matemáticos de los procesos, alimentados por enormes cantidades de datos de origen muy diverso, proporcionan resultados excepcionalmente precisos, pero, al mismo tiempo, ocultan la explicación de los procesos hasta el punto de llevarnos a pensar que la razón casi llega a ser innecesaria, siempre que las predicciones merezcan la pena.

El conocimiento de "qué sabemos" de la ontología es tan relevante en ciencia como el de "cómo sabemos" y el de "cuánto sabemos" de la epistemología; incluso en áreas de la investigación científica donde resalta el valor del cálculo numérico, el gran reto epistémico consiste en tratar de justificar el valor de unos procedimientos que, a partir de datos reales, permiten alcanzar veredictos predictivos o confirmatorios sobre la validez de una hipótesis, porque cuando la matemática se establece como una cárcel epistémica de la realidad pierde su valor.

En la ciencia real existe una fuerte interrelación entre evidencia, los valores epistémicos y no epistémicos, y los sesgos cognitivos. Con el fin de restaurar y promover la confianza pública en todos los ámbitos de la actividad científica, tal interacción debe hacerse explícita y discutirse públicamente; de hecho, las pruebas demostrativas son aceptadas según estándares científicos comunitarios y es precisamente la comunidad científica – en su amplia diversidad y libertad – la que proporciona confianza.

El desarrollo científico no es un modelo uniforme de evolución del conocimiento y, desde luego, no se debe exclusivamente al seguimiento de procesos racionales lógicos ni de rigurosos y precisos protocolos de metodología práctica. La intuición nos reta a intentar comprender por qué tiene un papel fundamental en el desarrollo de la ciencia, por qué permite alcanzar en muchos casos una conclusión mucho antes de disponer de pruebas objetivas que la avalen y profundizando más allá de la apariencia de lo evidente.

La inteligencia artificial (IA) implica la comprensión científica de los mecanismos que subyacen al pensamiento y la conducta inteligente, así como su encarnación en máquinas – agentes artificiales – capacitadas por sus creadores de razonar en un sentido convencional. Su formulación "débil" se refiere al empleo de programas informáticos complejos, diseñados con el fin de complementar o auxiliar el razonamiento humano para resolver o completar complejos problemas de cálculo, de mantenimiento de sistemas, de reconocimiento de todo tipo de

imágenes, de diseño, de análisis de patrones de datos, etc., muchos de los cuales serían prácticamente inabordables mediante procedimientos convencionales; pero todo ello sin incluir capacidades sentientes o éticas humanas, que sí serían objeto de una – por ahora – inexistente IA "fuerte", aquella que igualaría o incluso excedería la inteligencia sentiente humana.

La vulgarización de la IA "generativa", desarrollada para crear contenido – texto, imágenes, música o vídeos, entre otras muchas áreas – a partir de información previa, está contribuyendo a consolidar popularmente la idea errónea de que la actual IA excede el razonamiento a nivel humano y exacerba el riesgo de transmisión de información falsa y estereotipos negativos a las personas, algo que puede ocurrir incluso con del uso regular del sistema, en ausencia de fuerzas malévolas que busquen impulsar el sesgo o la desinformación.

Los modelos de lenguaje de la inteligencia artificial no funcionan emulando un cerebro biológico, sino que se fundamentan en la búsqueda de patrones lógicos a partir de grandes bases de datos procedentes de fuentes diversas, que no siempre actualizadas ni depuradas de falsedades, de errores ni de sesgos conceptuales o factuales, tanto involuntarios como interesados. Y la IA empleada en ciencia no es ajena a estas limitaciones y sesgos.

Otro de los problemas que pueden surgir de la IA no es que haga algo extraño o ajeno a las instrucciones de su programador, sino que haga lo que éste le dijo que tenía que hacer, pero sin que el programador fuese plenamente consciente del alcance y consecuencias de sus propias instrucciones. Esto en ciencia puede ser muy favorable, pero en tecnología – ciencia aplicada, en definitiva – no siempre tiene por qué serlo, incluso puede tener consecuencias catastróficas.

Son muy notables las posibilidades que abre la IA; sin embargo, no podemos olvidar que está diseñada por personas con manías, ideas, prejuicios, fobias y sentido existencial, y que están financiadas – y controladas estrechamente – por corporaciones privadas o públicas con intereses no siempre confesables.

Una cuestión particularmente sensible es la posibilidad de utilizar la IA generativa para redactar o incluso inventarse artículos científicos que llegan a pasar desapercibidos por los revisores por pares de las revistas científicas más prestigiosas del mundo. El uso no revelado de ChatGPT y otras herramientas de IA no solo se ha identificado en artículos de revistas científicas, sino también en conferencias revisados por pares y en preimpresiones. Sin embargo, el problema de los artículos no divulgados producidos por IA en revistas apunta a un problema más aún profundo: los revisores por pares de las revistas científicas a menudo no tienen tiempo para revisar los manuscritos a fondo en busca de señales de alerta y, en muchos casos, además carecen de recursos informáticos adecuados y formación especializada.

Por encima de todo, la ausencia de una voluntad libre para hacer algo es la gran limitación de la IA actual, que arrastra la doble condena de servir para un fin que ella misma no ha elegido y de no experimentar la alegría de conseguir sus metas ni la frustración de no haberlo hecho.

6. REFERENCIAS

1. Cuéllar Rodríguez S. Epistemología de la medición. An Real Acad Farm. 2022; 88 (1): 31-44. DOI: http://dx.doi.org/10.53519/analesranf.2022.88.01.02
2. Krakauer DC. At the limits of thought. Aeon. https://aeon.co/essays/will-brains-or-algorithms-rule-the-kingdom- of-science (2022)
3. Proctor RN. Agnotología (Agnotology). Revista de Economía Institucional, 2020; 22(42) Disponible en: https://ssrn.com/abstract=3495489
4. Amoretti MC, Lalumera E. COVID-19 as the underlying cause of death: disentangling facts and values. HPLS (History and Philosophy of the Life Sciences), 2021; 43: 4. https://doi.org/10.1007/s40656-020-00355-6.
5. Granville A. Why Mathematical Proof Is a Social Compact. Quanta Magazine, https://www.quantamagazine.org/why-mathematical- proof-is-a-social-compact-20230831/ 31 de agosto de 2023.
6. Fricker M. Epistemic injustice: power and the ethics of knowing.
Oxford: Oxford University Press. ISBN 9780198237907. (2007)
7. Cheng E. The joy of why: Is There Math Beyond the Equal Sign?
Quanta Magazine, https://www.quantamagazine.org/is-there- math-beyond-the-equal-sign-20230322/ 22 de marzo 2023.
8. Reeves RV. Lies and honest mistakes. Aeon. https://aeon.co/essays/our-epistemic-crisis-is-essentially-ethical- and-so-are-its-solutions (2021).
9. Cuéllar Rodríguez S. 239. Ediciones Vitruvio. ISBN: 978-84-949763-8-4 (2019)
10. Kidd C, Birhane A. How AI can distort human beliefs. Science. 2023
Jun 23; 380(6651): 1222-1223. doi: https://doi.org/10.1126/science.adi0248
11. Kwon D. Guillaume Cabanac: Deception sleuth. Nature, 15 december
2021; https://www.nature.com/immersive/d41586-021-03621-
0/index.html#section-gM9iO4XBRl:~:text=%3A

Indigenous defender-,Guillaume Cabanac%3A Deception sleuth,-Meaghan Kall%3A COVID
12. PubPeer. https://pubpeer.com/static/about
13. Conroy G. Scientific sleuths spot dishonest ChatGPT use in papers. Nature;08 September 2023 doi:https://doi.org/10.1038/d41586-023-02477-w.

Si desea citar nuestro artículo:
Epistemología y ontología en ciencia: el reto de la inteligencia artificial
Santiago Cuéllar Rodríguez
An Real Acad Farm (Internet].
An. Real Acad. Farm.Vol. 89. nº (2023) · pp. 379-386
DOI: http://dx.doi.org/10.53519/analesranf.2023.89.03.09