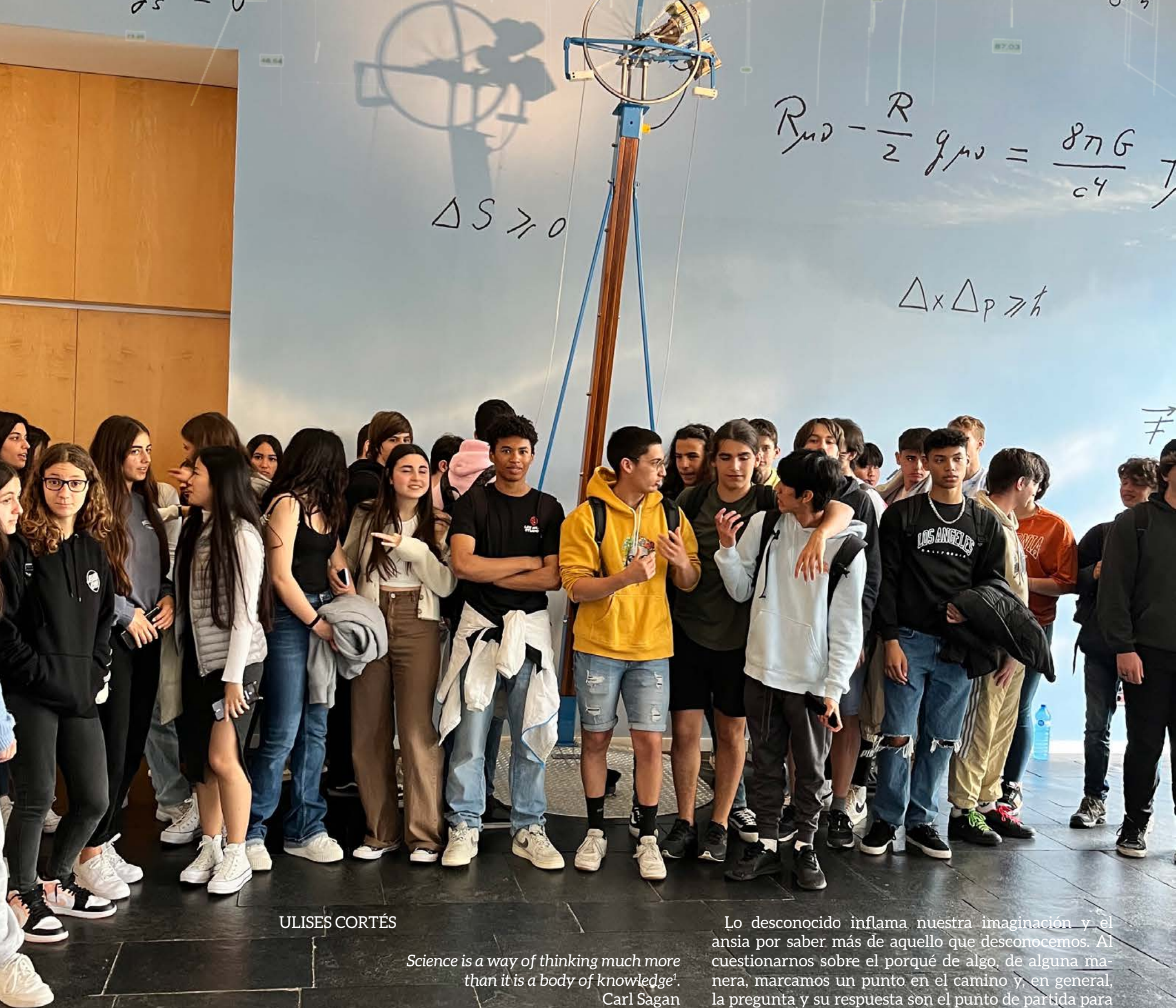


AI4Science



ULISES CORTÉS

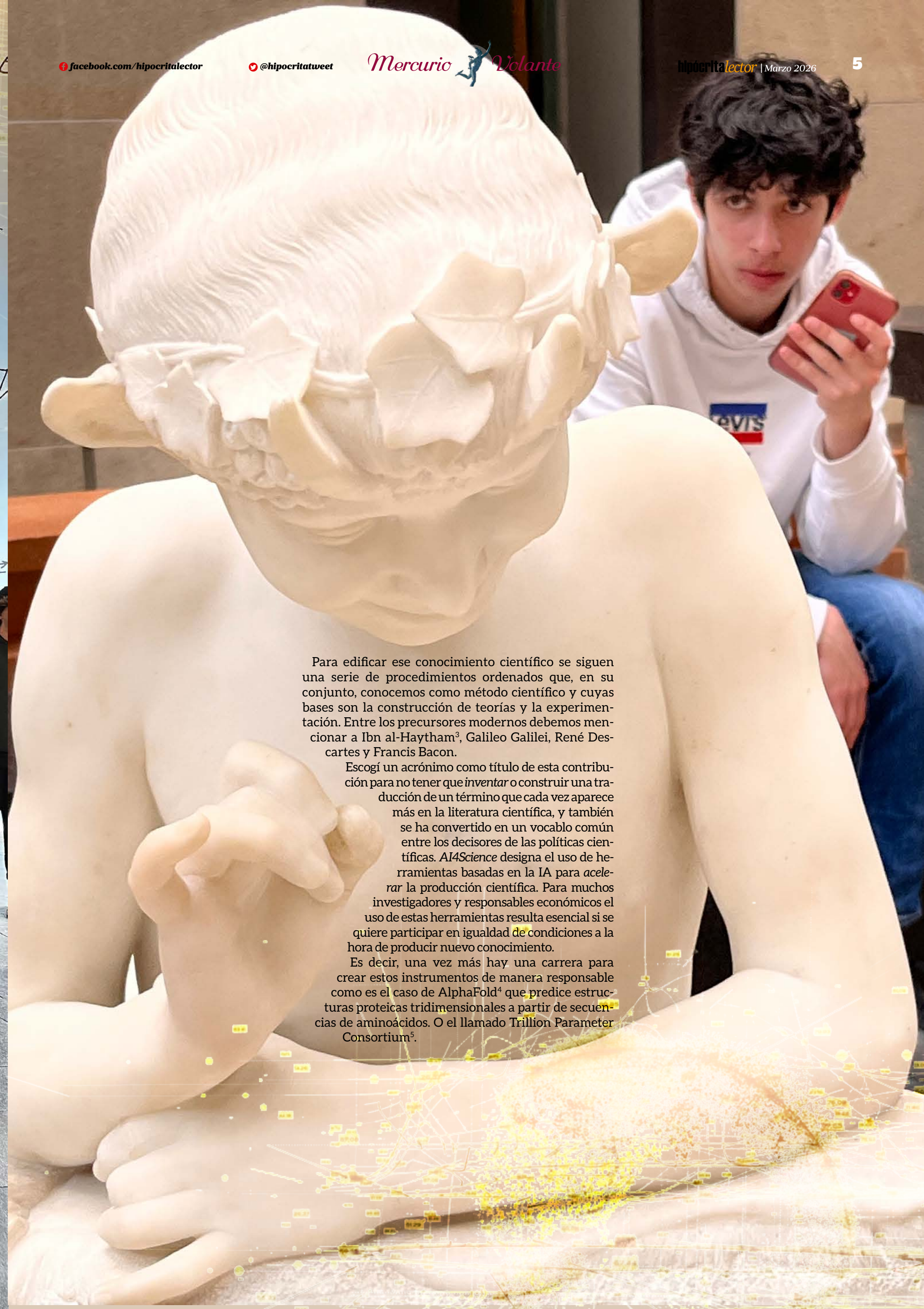
Science is a way of thinking much more than it is a body of knowledge.
Carl Sagan

Una traducción posible del epígrafe que acompaña a este artículo es *La ciencia es un cuerpo de conocimiento que ha sido recopilado a lo largo del tiempo; la ciencia es más mucho más que eso.*

Con esta definición tan abierta en mente, me gustaría aprovechar esta ocasión para invitarles a pensar más sobre la ciencia, a reconocer cómo ha moldeado la naturaleza de nuestra sociedad e incluso, cómo nos ha influido para hacernos pensar de cierta manera sobre las cosas que vemos, nos ha ayudado a identificar las que conocemos y, sobre todo, tratar de desentrañar, como apuntaba Carl Sagan, el misterio de las que no conocemos, desde las partículas subatómicas hasta las estrellas.

Lo desconocido inflama nuestra imaginación y el ansia por saber más de aquello que desconocemos. Al cuestionarnos sobre el porqué de algo, de alguna manera, marcamos un punto en el camino y, en general, la pregunta y su respuesta son el punto de partida para nuevas y más inquisiciones y sus respectivas respuestas. Requiere coraje y valentía hacerse nuevas preguntas o visitar las antiguas y sus resultados para avanzar en la búsqueda de nuevas piezas de conocimiento; pocas de estas respuestas son sencillas y cortas.

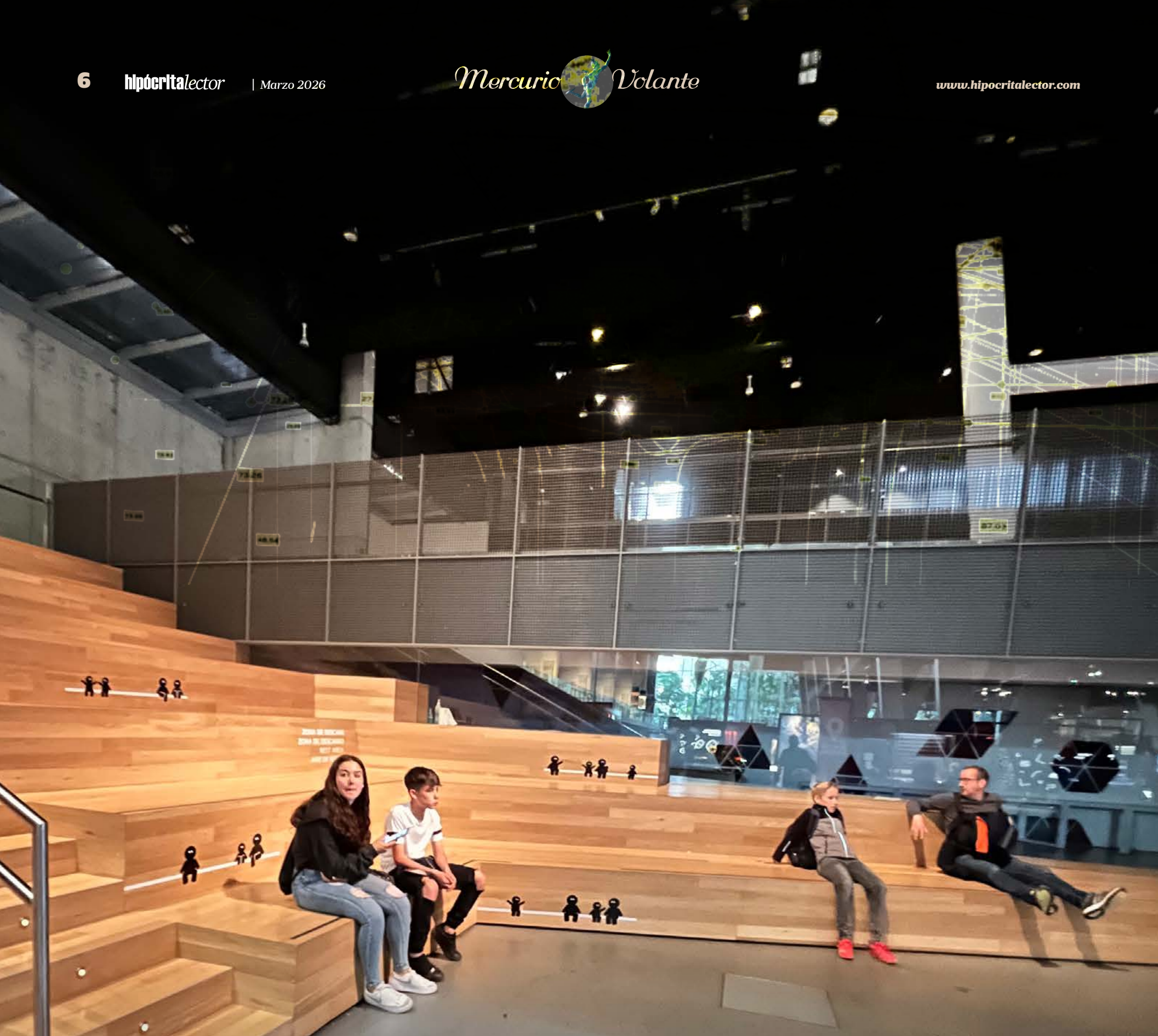
La ciencia, el cuerpo del conocimiento científico acumulado durante siglos, pasando de generación en generación, incluidos los conocimientos locales indígenas, aunque haya sufrido notables agresiones y retrocesos, es, sin duda, uno de los esfuerzos colectivos más prodigiosos y prolongados que la humanidad ha emprendido y, aunque solo recordamos a unos pocos de esos hombres y mujeres de los que llamamos científicos y su obra, su curiosidad y rigor son admirables; hoy, la perseverancia de muchos nuevos investigadores sostiene este proyecto común.



Para edificar ese conocimiento científico se siguen una serie de procedimientos ordenados que, en su conjunto, conocemos como método científico y cuyas bases son la construcción de teorías y la experimentación. Entre los precursores modernos debemos mencionar a Ibn al-Haytham³, Galileo Galilei, René Descartes y Francis Bacon.

Escogí un acrónimo como título de esta contribución para no tener que *inventar* o construir una traducción de un término que cada vez aparece más en la literatura científica, y también se ha convertido en un vocablo común entre los decisores de las políticas científicas. *AI4Science* designa el uso de herramientas basadas en la IA para *acelerar* la producción científica. Para muchos investigadores y responsables económicos el uso de estas herramientas resulta esencial si se quiere participar en igualdad de condiciones a la hora de producir nuevo conocimiento.

Es decir, una vez más hay una carrera para crear estos instrumentos de manera responsable como es el caso de AlphaFold⁴ que predice estructuras proteicas tridimensionales a partir de secuencias de aminoácidos. O el llamado Trillion Parameter Consortium⁵.



La calidad de estos sistemas se ha demostrado y reconocido al recibir Hassabis y Jumper la mitad del premio Nobel 2024 por el desarrollo de AlphaFold, Baker recibió la otra mitad por el diseño computacional de proteínas. Estas nuevas herramientas abarcan el espectro de toda la producción científica, lo que demuestra que AI4Science es un ecosistema en evolución, con una infraestructura propia y nuevas prácticas interdisciplinarias. Cada día se suman nuevas noticias de usos exitosos.

No obstante, en paralelo, se observa un uso acelerado, precipitado y poco crítico de estas tecnologías, en particular, de los modelos masivos del lenguaje⁶, sin una evaluación adecuada de sus consecuencias epistemológicas, éticas y sociales. Esta situación es preocupante en el ámbito académico y científico, donde los LLM se están incorporando a la redacción de artículos, la generación de hipótesis y el análisis de datos sin marcos de validación robustos ni protocolos claros de transparencia sobre su intervención en el proceso de investigación.

Cuando un autor envía un artículo para su evaluación y posterior publicación, presupone que hay un autor o autores que: ha(n) seleccionado las fuentes, los datos; ha(n) interpretado los resultados, ha(n) asumido una posición, y está(n) dispuesto a dialogar críticamente con los editores y los lectores. Esta falta de agencialidad es la característica

típica de los textos producidos de forma mecánica.

Sus defensores aducen que estas herramientas basadas en la IA al permitir un análisis de datos más eficiente, simulaciones potentes y nuevas formas de generación de hipótesis, ayudan a los investigadores a abordar problemas que antes eran demasiado complejos o requerían mucho tiempo para resolverlos.

Según sus defensores se disminuye el número de los experimentos necesarios o al menos los tiempos de experimentación; esta reducción afirman reduce los costes. Además, opinan que los LLM reducen los tiempos en el resumen de la literatura científica cuya abundancia y frecuencia de creación imposibilita estar al día o recordar artículos clásicos que de otra manera pasarían desapercibidos a la mayoría de los lectores. Aunque, dicho sea de paso, tampoco se leen dichos artículos.

Los efectos de esta adopción acelerada ya se manifiestan con claridad en múltiples disciplinas. En congresos y revistas científicas, por ejemplo, se percibe una creciente inflación de publicaciones, muchas de ellas de dudosa calidad o incluso abiertamente fraudulentas⁷. Hay evidencia de congresos científicos que de una a otra convocatoria anual han multiplicado por tres el número de propuestas de comunicación; esto bloquea los sistemas de revisión, pues no hay suficientes revisores humanos especialistas para juzgar esta avalancha.



● Estudiantes latinoamericanos en el Museo de Matemáticas en Manhattan, Nueva York.

Cuando se delegan tareas nucleares, como la redacción de secciones del artículo, la formulación de conclusiones o incluso la generación de datos sintéticos, en sistemas computacionales que no pueden asumir responsabilidad moral ni epistémica, se diluye la autoría científica y se vuelve opaco quién responde por los errores, sesgos o plagios que puedan derivarse.

Este fenómeno plantea interrogantes urgentes sobre la integridad de la producción científica, la fiabilidad de los procesos de revisión y la capacidad de las comunidades académicas para adaptarse al impacto de las herramientas de inteligencia artificial en la generación de información; queda abierta la discusión para saber si estas son creaciones científicas y si generan conocimiento o no.

Una respuesta a estos acontecimientos es la Colección de Guías de Integridad Científica Abierta (COSIG⁸) entre cuyos objetivos está el de capacitar a la comunidad científica para revisar de forma crítica la literatura ya publicada y detectar problemas de integridad, promoviendo una ciencia más transparente, verificable y fiable.

Ya hay quién plantea si habrá, gracias a esta tendencia, una ciencia sin científicos humanos. Llegados a este punto queda clara la necesidad de fortalecer el conocimiento científico entre la población a través de la educación. Hay que hacer llegar el mensaje de que con el pensamiento crítico y la búsqueda de la evidencia, tendremos más y mejor información para tomar mejores decisiones en nuestra vida cotidiana, pero también para enfrentarnos como sociedad a los grandes riesgos que la era antropocénica⁹ y del uso indiscriminado de las aplicaciones de la inteligencia artificial nos está trayendo.

Este modelo de aproximarse a la ciencia tiene riesgos que ni los científicos ni la sociedad parecen haberse planteado. Parece que la nueva forma de hacer ciencia consiste en hacer las preguntas adecuadas a un oráculo tipo LLM y se deja a las máquinas la tarea de responderlas. Detrás de los nuevos teoremas o las nuevas demostraciones ya no brillará el destello de un conjunto de neuronas humanas sincronizándose justo antes de exclamar ¡Eureka!

ULISES CORTÉS

Catedrático de Inteligencia Artificial de la Universitat Politècnica de Catalunya. Coordinador Científico del grupo High-Performance Artificial Intelligence del Barcelona Supercomputing Center. Miembro del Observatori d'Ètica en Intel·ligència Artificial de Catalunya y del Comitè d'Ètica de la Universitat Politècnica de Catalunya. Es miembro del comité ejecutivo de Eur AI. Participante como experto de México en el grupo de trabajo Data Governance de la Alianza Global para la Inteligencia Artificial (GPAI). Doctor Honoris Causa por la Universitat de Girona.



REFERENCIAS

- 1 Sagan, Carl (2011) [1979], *Broca's Brain: Reflections on the Romance of Science*, Random House Publishing Group, p. xiii, ISBN 9780307800992
- 2 Estos sistemas de conocimiento indígena han sido desdénados y olvidados de manera sistemática por la academia y la sociedad <https://www.unesco.org/es/links?hub=408>
- 3 [https://courier.unesco.org/en/articles/ibn-al-haythams-scientific-method#:~:text=Ibn%20al%20Haytham%20\(965%2D1040\)%20was%20a%2010th%2D11th%20century,and%20*On%20the%20Halo%20and%20the%20Rainbow*](https://courier.unesco.org/en/articles/ibn-al-haythams-scientific-method#:~:text=Ibn%20al%20Haytham%20(965%2D1040)%20was%20a%2010th%2D11th%20century,and%20*On%20the%20Halo%20and%20the%20Rainbow*)
- 4 <https://deepmind.google/science/alphafold/>
- 5 <https://www.anl.gov/cels/trillion-parameter-consortium>
- 6 Cabe recordar que los LLMs aparecieron hace solo tres años, esto es, el 30 de noviembre de 2022, y ya se han colado en casi todos los aspectos de la vida cotidiana <https://openai.com/index/chatgpt/>
- 7 Ozturk, Y., Bordignon, F., Labbé, C., & Portet, F. Assessing the Impact of Retracted Articles on the Reliability of Citing Literature: A Mixed-Based Proposal. In MET STI 2024. Zenodo. <https://hal.science/hal-04791200/>
- 8 <https://cosig.net>
- 9 <https://sciencemediacentre.es/antropoceno-que-es-y-por-que-crea-discusion-en-la-comunidad-cientifica>