La Inteligencia Artificial vista por un octogenario: de la regla de cálculo a los modelos de lenguaje

Jaime Terceiro Lomba
Real Academia de Ciencias Morales y Políticas
3 de noviembre de 2026

Resumen

En este ensayo analizo el progresivo cambio habido en la relación entre el ser humano y la máquina que ha desembocado en el salto cualitativo que supone la posibilidad de interactuar con ella en lenguaje natural e, incluso, delegarle tareas intelectuales complejas.

La introducción recorre mi propia experiencia, desde la regla de cálculo y la programación en Fortran hasta la mecanización de la administración y gestión de la Universidad Complutense, la modernización de los sistemas de información y decisión de un banco público, la enseñanza de métodos cuantitativos en economía en los cursos de licenciatura y doctorado y la mejora de la eficiencia operativa en una de las primeras instituciones financieras de España.

Un modelo de lenguaje es un sistema de inteligencia artificial (IA) que ha sido entrenado con ingentes cantidades de texto, y perfeccionado después mediante ejemplos y reiteradas valoraciones humanas, lo que le permite generar respuestas, tratar y resumir documentos, comparar distintos argumentos y proponer interpretaciones. Desde luego, estas capacidades no garantizan, por sí solas, la veracidad de sus respuestas ni implican una comprensión equivalente a la humana.

En la primera parte sitúo los modelos de lenguaje en su contexto histórico y técnico, y resume sus capacidades y límites. La segunda parte, que es el núcleo de la ponencia, contempla su posible utilidad en nuestra actividad académica y, concretamente, en la preparación, redacción y debate de nuestras ponencias semanales. La introducción y la primera parte son, en ese sentido, el contexto necesario; la segunda parte, el argumento.

Quedan fuera de este texto otros asuntos importantes y muy de actualidad. Por ejemplo, los efectos de la IA en el sistema educativo; la magnitud de las inversiones en este sector; su potencial rentabilidad y el riesgo de una burbuja financiera; su impacto en la productividad, el crecimiento y el mercado de trabajo; la carrera geopolítica por su dominio; su necesaria regulación; y, en definitiva, su repercusión en la sociedad y en la democracia. Cada uno de ellos merecería un

análisis propio y desviaría la atención del intento, sin duda mucho más modesto y restringido, de mejorar el contenido de nuestras sesiones de los martes.

Introducción

El título de esta ponencia pretende recordar el de Santiago Ramón y Cajal, *El mundo visto a los ochenta años* (1934), que a su vez evocó Alejandro Nieto —compañero que fue de esta Academia— con su libro *El mundo visto a los noventa años* (2022). A diferencia de ambos, yo no miro al mundo, sino solo a la IA y a lo que su aparición puede significar para nuestra actividad académica. Al motivo que me ha sugerido esta elección de título aludiré al final de la intervención.

Soy consciente de que traer a esta Academia una ponencia sobre IA requiere alguna justificación. No me extrañaría que alguno de ustedes sintiera ya un cierto hartazgo, teniendo en cuenta el enorme volumen de ensayos, artículos, opiniones e informes que este tema está suscitando en tan poco tiempo. Pues bien, mi justificación es doble: la primera es que no pretendo añadir un texto más a esa plétora, sino ofrecer una perspectiva propia cuyo origen explicaré enseguida; la segunda es que esa misma plétora forma parte del problema que quiero poner de manifiesto, pues cuando cualquiera puede obtener en minutos una síntesis aceptable de casi todo tipo de asuntos, lo apreciable realmente ya no es la información, sino el juicio necesario para ordenarla y evaluarla.

Debo pedir también tolerancia y comprensión por otro motivo. Una parte de esta ponencia la dedicaré a relatar mi relación con las máquinas, desde la regla de cálculo en 1964 hasta hoy. Me he atrevido a incorporar mi propia experiencia a este texto, aun a sabiendas de que hablar de uno mismo pone a prueba y tienta la paciencia de quien escucha o lee, sobre todo cuando el paso del tiempo hace que los recuerdos y la nostalgia tiendan a ocupar el lugar de los argumentos. Esta decisión responde a dos razones inherentes al propósito del trabajo, que no habría podido abordar de otro modo.

En primer lugar, de mi trayectoria personal concluyo que la IA representa una discontinuidad en la relación del hombre con la máquina tal y como yo la he vivido, y una discontinuidad solo se percibe con nitidez cuando se ha recorrido el progreso continuo que la precede; en segundo lugar, me ofrece la evidencia de que, al final de ese recorrido, la máquina podría absorber hoy mucho del trabajo que hice y muchas de las destrezas que durante décadas cultivé con mayor o menor fortuna. Esta constatación es la que me permitirá plantear en la segunda parte de esta ponencia la misma pregunta sobre nuestra propia actividad, por como la conozco desde que me incorporé a esta Academia hace ya treinta años.

Desde luego, no pretendo que mi trayectoria tenga valor o mérito particular alguno frente a la de tantos otros coetáneos que experimentaron una evolución similar. Por otra parte, es importante precisar que cuando aludo a máquinas me

estoy refiriendo concretamente a máquinas de calcular y de tratar información, es decir, a instrumentos destinados a mediar con la mente humana, en el manejo de símbolos y reglas con el propósito de solucionar problemas.

Hasta la llegada de los ordenadores personales a precios asequibles, solo dispuse en propiedad de dos de estos instrumentos: el primero fue una regla de cálculo Faber-Castell, adquirida en 1964, y el segundo una calculadora HP-21, en 1975. Los demás pertenecían a universidades, a centros de investigación o a empresas, y los utilizaba con arreglo a sus normas y su disponibilidad.

La primera etapa fue la de la regla de cálculo Faber-Castell, que empecé a utilizar durante el curso de ingreso de la Escuela Técnica Superior de Ingenieros Aeronáuticos. Era un instrumento analógico en su integridad, que por consiguiente no operaba con dígitos, sino que traducía los números en longitudes sobre escalas logarítmicas. Al deslizar la reglilla, sumar longitudes equivalía a multiplicar y restarlas equivalía a dividir. El usuario era el responsable de situar la coma decimal y decidir si la cifra obtenida tenía sentido. En definitiva, el acierto del cálculo dependía de su destreza y buen criterio.

La segunda etapa fue la del ordenador central. En 1970 realicé el proyecto de fin de carrera programando en Fortran, acrónimo de *formula translation*, en el Centro de Cálculo de la Universidad Complutense, que fue creado en 1966 mediante un acuerdo entre la Universidad e IBM. Su equipo incluía un IBM 7090 que ya había estado en funcionamiento en el CERN —Organización Europea para la Investigación Nuclear—, en Ginebra, y después en la Universidad de Heidelberg antes de llegar a Madrid, sobradamente amortizado.

Para trabajar con este ordenador el programa, o conjunto de instrucciones que se le daban a la máquina, se perforaba en tarjetas y, horas o días después de entregarlo a la lectora de tarjetas, se recogía el listado, que a menudo señalaba algún error en el código. Era lo que se llamaba trabajo por lotes. Cada iteración costaba tiempo, pero a la vez tenía una ventaja no buscada, ya que ese coste obligaba a la disciplina de tener que pensar con detenimiento el problema y el código correspondiente antes de dárselo al ordenador. Tengo que reconocer que ninguna tecnología posterior me ha vuelto a exigir semejante rigor antes de interactuar con una máquina.

La tercera etapa llegó con mi primera experiencia profesional, que comenzó en enero de 1971 en Messerschmitt-Bölkow-Blohm (MBB) cerca de Múnich. Su centro de cálculo combinaba infraestructuras analógicas y digitales muy avanzadas para la época. Disponía de ordenadores de la arquitectura IBM System/360 para el cálculo científico, el procesamiento de datos de ingeniería y la visualización gráfica; y también de miniordenadores Siemens de la serie 300 para el control de procesos, a los que llegamos a conectar periféricos de diseño propio para tratar flujos de datos en tiempo real. Aquella infraestructura industrial me

permitió, entre otras varias tareas, desarrollar aplicaciones de la formulación en espacio de los estados, hoy tan conocida y utilizada en economía cuantitativa, a plataformas inerciales y a una variedad de sistemas de control.

Por completar esta parte del relato, en 1975 dispuse del segundo de los instrumentos en propiedad, una calculadora científica de bolsillo no programable HP-21, ya digital. Utilizaba la notación polaca inversa, mostraba los resultados en una pantalla de LED rojos y, a diferencia de la regla de cálculo, se ocupaba ella misma de la coma decimal. El único vestigio mecánico que conservaba era el conmutador deslizante para seleccionar el modo de cálculo de funciones trigonométricas entre grados y radianes.

A comienzos de los años ochenta se abrió una cuarta etapa, en la que mi relación con las máquinas cambió de naturaleza. Ya no se trataba de utilizarlas para resolver solo mis propios problemas de cálculo, sino de decidir los procedimientos y técnicas con los que una institución pudiera resolver los suyos. Me había incorporado a la cátedra de Econometría y Métodos Estadísticos de la Universidad Complutense y, en enero de 1981, el rector Francisco Bustelo me nombró vicerrector primero, con responsabilidad sobre los asuntos económicos. La gestión de esa área —que incluía la presupuestaria, la política de compras y el seguimiento de las obras del campus, entre otras materias— se llevaba entonces prácticamente a mano. Para subsanarlo nos apoyamos en el centro de cálculo de la propia universidad, el mismo en el que una década antes había preparado mi proyecto de fin de carrera y que para entonces había mejorado considerablemente las prestaciones de sus equipos. En este caso el lenguaje que se utilizaba era básicamente Cobol, acrónimo de *Common Business-Oriented Language*, más adecuado para las aplicaciones de gestión empresarial e institucional. En un año quedaron «mecanizados», como se decía entonces, todos esos procesos.

Al cabo de ese año pasé al Banco Hipotecario de España, donde ocupé sucesivamente la Dirección General de Expansión y la de Inversiones. Allí encontré vigente, otra vez, un tratamiento manual tanto en aspectos críticos de la contabilidad general como en la gestión de los préstamos hipotecarios, que se consiguió mecanizar en un espacio de tiempo relativamente corto. Concluida esa labor, me incorporé con dedicación exclusiva a la universidad.

La quinta etapa comenzó en 1983, cuando asumí la dirección del Departamento de Economía Cuantitativa de la Universidad Complutense. La precariedad de medios limitaba la enseñanza econométrica a modelos de regresión con dos variables explicativas y muestras muy cortas, de modo que al estimarlos la matriz resultante, de orden 3×3, pudiera invertirse a mano en un tiempo razonable.

En un principio logramos la cesión al departamento de una máquina perforadora de tarjetas que los alumnos utilizaban para escribir sus códigos en Fortran, que era el lenguaje disponible que mejor se adaptaba a todo tipo de cálculo matemático. Una vez perforadas las llevaban en cajas, de autobús en autobús, desde Somosaguas hasta la calle Vitruvio, a más de quince kilómetros; días después recogían los resultados y, si había un error de sintaxis, debían repetir el proceso. Trece años después de mi proyecto de fin de carrera era, básicamente, el mismo trabajo por lotes.

Más adelante, conseguimos un terminal conectado con el Centro de Proceso de Datos del Ministerio de Educación y Ciencia que disponía de un ordenador central UNIVAC 1100. Así se logró eliminar buena parte del trasiego de cajas de cartón y acelerar considerablemente el trabajo. Recuerdo las laboriosas gestiones para poder conectar el terminal y obtener permiso para usarlo durante la primera quincena de cada mes, pues la segunda estaba reservada para elaborar las nóminas. Aun con esas limitaciones, fue un avance determinante, sobre todo para facilitar la formación de los alumnos.

En paralelo, continué adaptando el programa en Fortran que había desarrollado en MBB para estimar modelos econométricos en su formulación en espacio de los estados, y que ahora utilizaba con aquella infraestructura informática en los cursos avanzados de Econometría. Reescrito años más tarde en MATLAB, y conocido como E4, se convirtió en una herramienta general para especificar y estimar este tipo de modelos.

La última etapa llegó desde la dirección de Caja Madrid, cuya presidencia ejecutiva asumí en enero de 1988 y en la que estuve nueve años. La entidad disponía de una aceptable arquitectura informática, aunque relativamente centralizada, y una de mis primeras decisiones fue acometer su modernización y construir un nuevo centro de cálculo. El reto, en este caso, ya no era sustituir tareas manuales, sino decidir con qué entorno de tecnologías de la información y de la comunicación competiría en el mercado una entidad de esa dimensión a finales del pasado siglo.

Para entonces habían llegado ya los ordenadores personales. Por primera vez pusieron a mi alcance, en propiedad y a un precio asequible, una capacidad de cálculo programable. Con ellos también aparecieron lenguajes próximos a la notación matemática, como MATLAB, acrónimo de *MATrix LABoratory*. Así pues, el entorno pasó a ser interactivo y la programación se aproximó a la forma en que por entonces ya desarrollaba mi actividad como profesor de métodos cuantitativos en economía.

Todo este recorrido permite formular la conclusión que anuncié al principio de la ponencia. Cada etapa fue reduciendo la distancia entre el pensamiento humano y su ejecución por la máquina. Utilizar la regla de cálculo requería

destreza y juicio; las tarjetas perforadas, precisión formal; MATLAB permitió que el diálogo fuera inmediato, pero todavía en un lenguaje formal. La evolución que llevó de Fortran a MATLAB, a herramientas como SAS, GAUSS, TSP o el propio E4 y, más tarde, a lenguajes que ya no llegué a utilizar, como Python —cuyo nombre no es un acrónimo, sino un homenaje al grupo de humoristas británicos *Monty Python*— o Julia, fue iterativa y gradual. Cada nueva herramienta hacía más amable la relación con la máquina, pero la naturaleza del diálogo seguía exigiendo un lenguaje formal. Pues bien, la IA generativa —capaz de elaborar nuevos textos, imágenes o código, y cuyo principal ejemplo son los modelos de lenguaje— al aceptar el lenguaje natural introduce una tajante discontinuidad. A mi juicio, ese cambio de interfaz, más que un aumento aislado de potencia de cálculo, es el verdadero punto de ruptura.

1. La máquina que aprendió a hablar con nosotros

La evolución que he relatado en primera persona coincide con la historia general de las interacciones con los equipos informáticos, y es también la de una progresiva ampliación del acceso al cálculo y al tratamiento de la información. El uso de tarjetas perforadas requería formación especializada y acceso a centros de cálculo que eran escasos; los lenguajes de alto nivel, así como las ventanas, el ratón y las pantallas táctiles, fueron reduciendo paulatinamente esas barreras. Los modelos de lenguaje —que se entrenan con grandes cantidades de texto— han supuesto un gran paso adelante porque el sistema puede interpretar instrucciones directamente en lenguaje natural, evitando al usuario la necesidad de dominar una sintaxis de lenguaje formal. Este cambio tiene drásticas consecuencias institucionales, económicas y académicas.

El ecosistema de modelos de lenguaje es competitivo e inestable. Los grandes sistemas generalistas, como GPT, de OpenAI, Claude, de Anthropic, Gemini, de Google, Grok, de xAI, los desarrollados por Meta o los chinos DeepSeek y Qwen, este último de Alibaba, se renuevan cada pocos meses. A la vez proliferan herramientas especializadas en campos tan diversos como la investigación y la enseñanza, la administración, el derecho, la medicina y las finanzas. Esta tecnología se caracteriza tanto por su potencia de cálculo como por su accesibilidad, puesto que hoy en día muchas de estas capacidades están al alcance de cualquier persona que disponga de conexión a internet, si bien puede haber diferencias dependiendo del tipo de servicio y sus condiciones de acceso.

1.1 De Dartmouth a la IA de hoy

Cuando en 1964 empecé a utilizar la regla de cálculo, la disciplina que hoy llamamos IA llevaba ya ocho años de existencia. Había nacido formalmente en el verano de 1956, cuando John McCarthy, un joven profesor de matemáticas, reunió en el Dartmouth College a un grupo de investigadores para estudiar si las

máquinas podían llegar a realizar tareas propias de la inteligencia humana. La propuesta con que se convocó aquel encuentro, formulada en 1955 junto con Marvin Minsky, Nathaniel Rochester y Claude Shannon, acuñó el propio término *inteligencia artificial*, véase McCarthy *et al*. (1955). Así que mi trayectoria personal con las máquinas de calcular arrancó cuando la disciplina daba sus primeros pasos, pero, obvio es decirlo, por una vía completamente independiente, pues ni la regla de cálculo debía nada a Dartmouth, ni en Dartmouth se había pensado en reglas de cálculo.

Hay que señalar, sin embargo, que la cuestión de fondo respecto a la factibilidad de una inteligencia artificial era anterior. Ya en 1950, Alan Turing se había preguntado «¿pueden pensar las máquinas?». Después reemplazó esa cuestión por otra más operativa, ¿podría una máquina conversar por escrito con una persona sin que esta lograse distinguirla de un interlocutor humano?, véase Turing (1950). Que el criterio elegido fuera, precisamente, el dominio del lenguaje natural resulta muy significativo.

En aquellos primeros años lo que predominó fue una concepción de la inteligencia basada en reglas que consistía en indicar a las máquinas cómo razonar y resolver problemas mediante instrucciones diseñadas por los propios investigadores. Sin embargo, y paralelamente, se retomó otra idea con origen en los años cuarenta, de que bastaba con proporcionar a la máquina ejemplos y dejar que aprendiera de ellos, en lugar de dictarle cada regla específica; esta corriente, conocida entonces como *reconocimiento de patrones* (*pattern recognition*), acabaría teniendo una importancia decisiva. Así, en 1958, Frank Rosenblatt desarrolló en el Laboratorio Aeronáutico de Cornell el perceptrón, que fue uno de los primeros modelos de ese tipo, capaz de aprender a distinguir entre diferentes clases de imágenes, véase Rosenblatt (1958).

Conviene precisar que *inteligencia artificial* y *reconocimiento de patrones* no son términos intercambiables, aunque a veces se empleen como si lo fueran. *Inteligencia artificial* hace referencia al objetivo general de todo el campo, es decir, que una máquina se comporte de forma inteligente sea cual sea la vía que utilice. En cambio, *reconocimiento de patrones* solo se refiere a la vía que consiste en proporcionar al sistema ejemplos a partir de los cuales aprende regularidades; con el paso del tiempo confluyó con el *aprendizaje automático* (*machine learning*). La corriente basada en reglas y la del aprendizaje a partir de ejemplos siguieron durante décadas caminos separados.

La evolución de la disciplina fue compleja y desigual. Con la publicación de su trabajo Minsky y Papert (1969) evidenciaron importantes limitaciones de los perceptrones. De hecho, sus críticas alimentaron la controversia sobre las posibilidades reales de estas técnicas y, aunque no fueron la única causa, contribuyeron al estancamiento de la investigación en este campo. Poco tiempo después, las excesivas expectativas depositadas en la IA provocaron serios

problemas de financiación. Por ejemplo, el informe Lighthill (1973) tuvo especial impacto en el Reino Unido y propició los recortes económicos que caracterizaron el que se conoce como *primer invierno de la IA*.

A finales de los años ochenta la confianza depositada en los sistemas expertos —programas que codificaban en reglas el conocimiento de los especialistas— se topó con sus limitaciones y, adicionalmente, con los elevados costes de su desarrollo y mantenimiento. Empezó entonces el *segundo invierno de la IA*.

El avance de la IA recibió un impulso decisivo gracias a la concurrencia del aumento de la capacidad de cálculo, de la creciente disponibilidad de información digitalizada y de la aparición de métodos de aprendizaje más eficaces. Internet no fue un factor ajeno a este proceso al facilitar el acceso a enormes cantidades de textos, imágenes y todo tipo de datos.

Uno de los métodos que hizo posible este avance fueron las redes neuronales artificiales, sistemas inspirados en el funcionamiento del cerebro, y que aprenden a partir de ejemplos. Están formadas por unidades conectadas entre sí y distribuidas en distintas capas que ajustan sus conexiones durante el entrenamiento. De este modo, pueden identificar patrones y realizar tareas sin necesidad de programación previa. Toda red neuronal tiene una capa de entrada y otra de salida, y cuando entre las dos hay varias capas intermedias, se dice que es profunda. El perceptrón de Rosenblatt, por ejemplo, no lo era; pero los modelos de lenguaje actuales, a los que me referiré en el siguiente epígrafe, sí lo son.

En 2012 se produjo un momento clave cuando una red neuronal profunda llamada AlexNet logró un resultado espectacular participando en una competición internacional en la que los sistemas informáticos debían reconocer y clasificar imágenes. La competición utilizaba una selección de imágenes de ImageNet, una extensa base de datos creada para impulsar el desarrollo de la IA en el reconocimiento de imágenes.

AlexNet utilizaba numerosas capas de procesamiento para aprender a identificar particularidades de las imágenes a partir de un gran número de ejemplos. Su éxito puso de manifiesto las posibilidades de combinar estas redes con abundantes datos y con una elevada capacidad de cálculo, y marcó un hito en el desarrollo de la IA; véase Krizhevsky, Sutskever y Hinton (2012). Desde entonces, el aprendizaje automático y, en particular, el aprendizaje profundo (*deep learning*) han pasado a ocupar un lugar central en la IA.

Queda, sin embargo, la incógnita de que, en caso de no llegar a cumplirse las enormes expectativas depositadas hoy en la IA, podría abrirse un nuevo ciclo de decepción, o incluso de fracaso, análogo a los anteriores. Lo que daría paso al

que sería el *tercer invierno de la IA*, aunque no hay acuerdo sobre si llegará ni sobre cuándo podría producirse. Ahora bien, las consecuencias de todo tipo de este posible *tercer invierno* serían mucho más severas que las de los dos primeros, dado que la IA ya no es un mero experimento académico.

1.2 Entrenamiento, capacidades y límites

Para poder aprender las regularidades del lenguaje y de la forma en la que las personas expresan sus conocimientos, los modelos se entrenan con enormes cantidades de texto y en esta tarea ajustan miles de millones de parámetros. Después se perfeccionan con ejemplos de respuestas adecuadas y con la evaluación humana de sus resultados. La arquitectura Transformer, véase Vaswani et al. (2017), mejoró la capacidad de los modelos para relacionar distintas partes de textos, incluso cuando se encuentran alejadas. Este avance contribuyó decisivamente al desarrollo reciente de los modelos de lenguaje.

Conviene distinguir entre el modelo de lenguaje y la aplicación que permite utilizarlo. Por ejemplo, ChatGPT y Claude son los nombres de las aplicaciones de OpenAI y Anthropic, respectivamente, aunque Claude también da nombre a los modelos de lenguaje de esta última compañía. Ambas son herramientas que dan acceso a dichos modelos y que pueden incorporar funciones adicionales, como la búsqueda de información o la consulta de documentos. En lo que sigue, me referiré principalmente a las capacidades de los modelos de lenguaje tal y como las he experimentado mediante estas aplicaciones.

Sus prestaciones son hoy muy amplias; sin embargo, el hecho de que sus respuestas sean fluidas no implica que sean necesariamente correctas. Estos modelos pueden hacer afirmaciones aparentemente creíbles que no siempre son ciertas. Por ejemplo, en la ayuda a un trabajo académico, pueden facilitar citas incorrectas, simplificar teorías o presentar como ya consolidada una conclusión todavía discutida. No verifican por sí solos los hechos y cuando establecen relaciones causales no lo hacen con garantía; véase Chi *et al.* (2024). Por tanto, el usuario siempre debe comprobar cuándo una respuesta está fundada y cuándo solo parece convincente, así como contrastar las fuentes que utiliza.

Los modelos aprenden, en buena medida, de textos digitalizados que están disponibles para su entrenamiento. Sin embargo, en muchas ocasiones el acceso a la información es limitado y desigual, como ocurre con las revistas de suscripción, libros enteros, archivos institucionales y datos todavía no publicados. Aunque pueden consultar internet para generar la respuesta, eso no elimina las barreras de acceso. Además, su entrenamiento se cierra en una fecha determinada, por lo que no pueden incorporar información y acontecimientos posteriores. Por todo ello, conocen deficientemente la literatura especializada más actual y tienden a reproducir con facilidad lo que más circula, que no siempre es lo más relevante.

Cabe preguntarse, entonces, por qué tienen interés los resultados de un sistema que carece de tanta información, buena parte de la cual es la más reciente. A mi juicio, por tres razones.

La primera razón de su utilidad es que no se limitan a reproducir información, sino que también pueden reconocer formas de plantear problemas, desarrollar argumentos, formular objeciones y traducir las ideas en código. Por eso conviene distinguir entre qué datos tienen a su alcance y qué operaciones pueden realizar con ellos.

La segunda es que parte de la información contenida en obras de acceso restringido puede estar indirectamente accesible, por ejemplo, en reseñas, citas, resúmenes y otro tipo de trabajos que el modelo sí pueda consultar. No obstante, en ese tránsito se pierden profundidad y detalle, y es ahí donde resulta imprescindible el saber del especialista humano y la consulta de las obras originales.

La tercera, y la más práctica, es que muchas funciones dependen del material que aporta el propio usuario. Si se le facilita un trabajo, un conjunto de datos o un simple esquema, el modelo puede analizarlo sin que ese material haya formado parte de su entrenamiento previo.

En todo caso, su utilidad es real, pero condicional. De ahí la regla que guiará el resto del trabajo, según la cual la IA propone, pero el experto verifica y responde por lo que finalmente sostiene.

Recientemente se han desarrollado sistemas de IA para actuar como agentes que pueden encadenar tareas con creciente autonomía, desde la búsqueda de información a la redacción de textos. No obstante, que una persona fije el objetivo no garantiza por sí solo una supervisión eficaz, y por ello sigue siendo necesario controlar y revisar los resultados, que deben quedar siempre bajo responsabilidad humana.

2. La IA y nuestra actividad académica

Esta es la primera ponencia en la que he utilizado la IA como instrumento de trabajo. Me ha ayudado a revisar intervenciones de los tres últimos años de esta Academia, sus turnos de preguntas, y a preparar un esquema preliminar de este texto. Me parece obligado explicitarlo, como por otra parte se hace en cualquier trabajo en el que no solo se reconocen las fuentes, sino también a quienes han contribuido a él de alguna forma.

Pienso que el verdadero valor de las ponencias que en este pleno debatimos es el de contener una perspectiva genuinamente propia sobre el tema que se trata. No me refiero, desde luego, a tener que traer aquí resultados de investigación originales publicables en revistas de prestigio. Conviene distinguir

esta originalidad teórica o empírica, que aporta teorías o resultados nuevos, de la originalidad interpretativa, que ofrece una lectura propia y bien fundada sobre el tema que se expone. Creo que es esta segunda la que aquí importa, y que ha nacido a lo largo de años de lectura, investigación o de determinadas experiencias tanto en el sector público como privado. Quien ha dedicado décadas a estudiar o a gestionar en un determinado ámbito tiene una comprensión e interpretación sobre todo ello que no está en lo publicado, sino probablemente en lo que ha descartado y en el conocimiento que orienta su lectura de los viejos y nuevos problemas. Además, permite ponderar la importancia de cada dato en la valoración de un problema, distinguiendo interpretaciones superficiales y reconociendo cuándo un debate actual repite controversias del pasado.

2.1 Un precedente ya vivido

En la introducción adelanté que, al final de aquel recorrido sobre mi relación con las sucesivas generaciones de instrumentos de cálculo, la máquina podría absorber hoy buena parte del trabajo que hice y de las destrezas que intenté cultivar durante décadas.

Esta situación es perfectamente aplicable a la de muchos profesionales y académicos. Probablemente, el ejemplo más claro sea el del programador como la persona que trasladaba con rigor, después de haberlo interpretado adecuadamente, un problema matemático a instrucciones y operaciones ejecutables por una máquina. Era esta una competencia relativamente escasa entonces, y que en gran parte puede ser desempeñada en la actualidad por la IA. No obstante, lo que todavía no ha absorbido la máquina, y que sí poseían los programadores mejor formados de aquella época, es la comprensión de por qué un método, expresado en un determinado conjunto de ecuaciones, resulta adecuado, qué supuestos lo sostienen y cuáles son las consecuencias de su incumplimiento.

Lo más instructivo de este proceso no es el reemplazo por la máquina de determinadas actividades humanas, sino el modo en que respondieron a él quienes lo vivieron. Unos se aferraron a la destreza, o a parte de ella, que dominaban y perdieron su ventaja relativa cuando la máquina la hizo innecesaria; sin embargo, otros reaccionaron desplazando su esfuerzo hacia la comprensión de aquellos problemas y de otros nuevos, y salieron reforzados, porque esta es precisamente lo que la automatización todavía no alcanza.

Sería un error, no obstante, interpretar lo anterior como un juicio sobre lo ya hecho. La actividad en mis primeros años en MBB, la docencia y la investigación que desarrollábamos más tarde en el Departamento de Economía Cuantitativa, o la modernización de los sistemas de información y decisión de Caja Madrid no podrían repetirse hoy de la misma manera, ni sería razonable intentarlo, porque han cambiado las condiciones de contorno. Pero que no deban repetirse no les

resta validez. Prueba de ello es que los alumnos que se formaron con un material informático tan precario han llegado a ocupar puestos muy relevantes en instituciones y universidades de España, de la Unión Europea y de EE. UU. Y el proceso de modernización de Caja Madrid contribuyó a situarla entre las entidades financieras españolas más rentables y solventes. Hay que recordar que programando en Fortran y perforando tarjetas se calcularon en buena parte las trayectorias que llevaron al hombre a la Luna. Aquellos trabajos fueron los adecuados a su tiempo y son la base para entender hoy los sistemas y actividades que se automatizan; lo que ha cambiado no es el mayor o menor mérito que en su día se les reconoció, sino las circunstancias en las que se desarrollaron.

De lo dicho hasta aquí el paralelismo que podemos establecer con la actividad de nuestra Academia surge de manera natural. Desde mi incorporación a esta casa pienso que la forma y el fondo con el que hemos venido preparando, presentando y discutiendo las ponencias en estas sesiones apenas han cambiado. Nada en las condiciones de trabajo en las que desarrollamos esta tarea lo ha propiciado. Algunas de las razones endógenas para esta ausencia de cambio las expondré más adelante. Sin embargo, tenemos que ser conscientes de que, como ya he señalado, la accesibilidad al ecosistema de los modelos de lenguaje está teniendo un considerable impacto en multitud de actividades y, especialmente, en las relacionadas con el ámbito de la academia y de la investigación. En efecto, una vez que elegimos un tema de interés estas herramientas permiten acceder a buena parte de la literatura de referencia y compendiarla. Y planteado ya un esquema de trabajo o un borrador de texto son de gran ayuda para identificar inconsistencias y contradicciones, para proponer objeciones y para sugerir posibles desarrollos alternativos.

De todas estas funciones, en mi opinión la más valiosa es la de abogado del diablo. Bien utilizada, y después de un cierto periodo de práctica, la IA puede servir al ponente como un ayudante de investigación que, interrogado adecuadamente, rebate nuestros argumentos desde perspectivas distintas y señala afirmaciones excesivas o falta de pruebas. No siempre advertirá por iniciativa propia la objeción decisiva, e invariablemente es necesario reformular las preguntas desde ángulos diferentes. Este proceso exige un constante intercambio de opiniones y sugerencias con este singular «ayudante de investigación».

Cabe entonces que nos hagamos el mismo tipo de pregunta que, por ejemplo, se hizo el programador, ¿qué parte de nuestra actividad es mera implementación y ordenación de ideas ajenas, asumible cada vez más por un modelo de lenguaje, y qué parte es comprensión y juicio propio? Dicho en otros términos, en una ponencia lo que aquí llamo implementación consiste en ordenar y, en su caso, resumir lo que otros han dicho, mientras que el valor añadido

propio está en dar una nueva interpretación a todo este material y en hacer explícitas las razones por las que el autor la defiende.

2.2 Juicio, coherencia y erudición

Lo que desde luego la IA no puede sustituir es la combinación de juicio, responsabilidad y confianza del ser humano. Por ejemplo, el juicio bajo incertidumbre, en el sentido de Frank H. Knight (1921), que la distingue del riesgo al que puede asignarse con rigor una determinada probabilidad, exige ponderar una serie de factores y sus consecuencias. La responsabilidad frente a este pleno, frente a un consejo de administración o a un regulador no se puede delegar en algoritmo o modelo alguno. Lo mismo sucede con la confianza. Así, las instituciones, y valga como ejemplo esta Academia entre ellas, funcionan porque sus miembros confían unos en otros, en su palabra y en su criterio. Esta confianza no surge espontáneamente, sino que se cimienta con la conducta a lo largo del tiempo y descansa en que las personas responden de sus actos. Por estas razones no cabe depositarla en un modelo, por muy útil y eficiente que resulte en otras tareas.

Quien ha dirigido instituciones lo sabe por experiencia. Si hoy se sometiera a la dirección de una entidad financiera la aprobación de un modelo de IA para conceder crédito, habría que aclarar qué decisiones toma el sistema y cuáles se reservan a una persona o a un departamento identificable, quién entiende el modelo lo suficiente para cuestionarlo y si el hecho de que el conjunto del sistema emplee el mismo modelo puede generar correlación en los errores del proceso de concesión de créditos por las demás entidades. Saber distinguir cuándo un modelo representa la realidad y cuándo la desfigura es una forma de experiencia que merece convertirse en argumento explícito.

Traigo este ejemplo a colación no solo por haber vivido situaciones semejantes, sino también para recordar que el Reglamento de IA de la Unión Europea considera de alto riesgo los sistemas que evalúan la solvencia de las personas para concederles crédito, y les impone, entre otras obligaciones, la supervisión humana, que será exigible a partir del 2 de diciembre de 2027. Aunque, como ya he señalado, en este trabajo no abordo la regulación de la IA, quiero aprovechar la ocasión y el ejemplo para expresar mi decidida posición a favor de regular esta industria. Los economistas sabemos que el poder de mercado, los diversos tipos de externalidades y las asimetrías de información justifican en este caso la intervención pública, como en su día la justificaron en la industria farmacéutica, la aviación civil o la energía nuclear, cuya regulación tuvo éxito sin ir en detrimento de su desarrollo.

En nuestro ámbito académico, el juicio propio, o lo que llamé originalidad interpretativa, se sustituye con frecuencia por la erudición. Cuando este es el caso aparecen debilidades propias de esta manera de proceder. Las he agrupado en

tres categorías distintas, y a ellas me voy a referir a continuación de menor a mayor dificultad para percibirlas, aunque pienso que la segunda es la más grave.

La primera es la ausencia de una tesis propia. A pesar de la sólida erudición, no aparece explícitamente en el texto una posición que se sostenga con claridad y por la que el autor responda. Si bien la exposición informa y puede aportar datos, por sí sola no llega a constituir un argumento. Con frecuencia, en estas situaciones se desarrolla una analogía histórica conocida, sin poner de manifiesto por qué resulta hoy pertinente.

Una variante de esta categoría es aquella que afronta un gran tema, generalmente de actualidad, que ya ha sido abordado muchas otras veces, por personas e instituciones muy capaces desde múltiples puntos de vista, y en la que se ordena con más o menos oficio lo ya dicho por otros para concluir con una reflexión general que nadie podría discutir. Estos casos no son simples ejercicios de erudición, ya que contienen una posición, pero es tan amplia que no compromete al autor ni admite refutación alguna. Por otra parte, esta es la clase de trabajo que un modelo de lenguaje produce con mayor facilidad, porque apenas hace otra cosa que reunir lo que ya circula.

La segunda categoría es la que hace referencia a las debilidades de fundamento. Una tesis propia puede apoyarse en argumentos teóricos equivocados o no aplicables al caso, o en evidencia empírica tratada inadecuadamente. Que una conclusión sea plausible, y con apariencia de sensatez, no vuelve correcto el razonamiento que conduce a ella. Puede incluso que la conclusión sea acertada, pero que se le atribuyan causas equivocadas. Por eso es, a mi juicio, la categoría más grave, ya que nadie discute una conclusión que le parece correcta, o incluso es correcta, y el razonamiento espurio circula y se repite sin ser advertido. En estas situaciones poco puede aportar la IA, ya que, aunque ayuda a detectar inconsistencias lógicas, no está en condiciones, como ya se ha señalado, de establecer causalidad. En este ámbito la econometría ha hecho progresos muy sustanciales en las últimas décadas sin necesidad de recurrir siempre a experimentos controlados. Valga como ejemplo el Nobel de Economía de 2021, concedido a Angrist e Imbens por sus contribuciones metodológicas a este problema y a Card por sus aportaciones empíricas a la economía laboral. Cuando no hay una evidencia empírica sólida que permita establecer una determinada relación causal, la respuesta rigurosa es reconocerlo.

En esta categoría sitúo también el uso de la cita de autoridad como alternativa al análisis riguroso. Invocar a un clásico no refuta por sí solo un argumento contemporáneo que aquel nunca pudo conocer. Una variante de este caso es seleccionar de una obra de larga tradición interpretativa una cita favorable a casi cualquier posición; el *Quijote* es el ejemplo paradigmático. No obstante, elegir entre pasajes contradictorios no es de ayuda alguna, ya que es exigible justificar cuál de las lecturas es válida para el problema entre manos. También

aquí cabe incluir el caso de textos que habiendo resumido ideas ya publicadas lo hacen citando a autores periféricos y pasan por alto a aquellos que son centrales a las ideas glosadas y en los que realmente se apoyan.

La tercera categoría abarca los fallos en la arquitectura argumental, y con frecuencia es la más difícil de advertir, ya que está cubierta de una apariencia de rigor. Es el caso de la yuxtaposición de citas o marcos teóricos incompatibles entre sí, construidos por acumulación bibliográfica, invocando a una cadena de autores sin reparar en las contradicciones entre ellos. Tal defecto no supone, necesariamente, una falta de suficiente conocimiento por parte del autor, sino la ausencia de una tesis central que sirva para ordenar las referencias. En mi experiencia un modelo de lenguaje puede detectar estas contradicciones que en una redacción o lectura humana suelen pasar desapercibidas. Ahora bien, poner de manifiesto este tipo de incoherencias no resuelve el problema ya que su necesaria valoración corresponde al autor.

Otra situación que está en esta categoría es la falta de jerarquía. No basta con enumerar y documentar los problemas con esmero; es necesario priorizarlos, sopesar su nivel de incertidumbre y proponer la pertinente intervención. Un diagnóstico sin jerarquía no es todavía un diagnóstico, sino una lista de problemas potenciales, y cuanto más extensa es esa lista menor es la probabilidad de equivocarse en el diagnóstico, precisamente porque nada afirma sobre qué problemas son los más relevantes. La IA produce inventarios con facilidad, pero para que sean útiles se necesita el criterio humano.

A este análisis añade la IA un elemento nuevo que no crea el problema, pero sí lo pone dolorosamente de manifiesto. Si buena parte de lo que se publica en ciencias sociales consiste en reordenar con habilidad argumentos y lecturas ya conocidos, que es la implementación intelectual de la que vengo hablando, un modelo de lenguaje actual puede producir en minutos, y con un mínimo de instrucciones, una síntesis de calidad comparable, en mi opinión, a la de no pocos ensayos que llegan a ver la luz tras meses o años de trabajo. Desde luego, esto no pone en entredicho al autor cuando aporta un juicio propio que la máquina no tiene, pero sí descalifica con nitidez a la parte de la producción académica que nunca aportó tal juicio y que solo ahora, por comparación, se revela como redundante. La pregunta que cada autor debería hacerse antes de publicar es la de si una IA diría lo mismo igual de bien con la misma información. Es esta una pregunta incómoda, aunque no por ello menos pertinente. Si la respuesta es afirmativa y el texto no aporta una contribución propia, en el sentido, al menos, de la originalidad interpretativa antes descrita, la publicación puede añadir algo a determinado tipo de currículum, pero no añade conocimiento.

La situación encierra, además, una doble paradoja. Al abaratar la síntesis, la exposición y la revisión bibliográfica, la IA eleva el valor relativo de los trabajos verdaderamente originales, porque lo que pasa a escasear es el hecho de tener

algo importante o significativo que decir. Pero, a la vez, la misma IA que hace innecesarios muchos textos que hoy se publican facilita también la producción de muchos más. La oferta de obras de escaso valor intelectual podría crecer al tiempo que se pone de manifiesto su inutilidad.

2.3 Turno de preguntas

El turno de preguntas debería permitir comprobar si la ponencia contiene un juicio propio. En la práctica las objeciones pueden moderarse por razones ajenas a la capacidad intelectual de los asistentes, como la cortesía, la reciprocidad o la incomodidad de criticar públicamente a colegas. Pues bien, al preparar este trabajo y revisar otros míos anteriores, constaté que algunas observaciones realizadas por un modelo de lenguaje me resultaron más precisas e incómodas que las de un turno de preguntas ordinario. Esto sucede no porque la máquina sea necesariamente más inteligente, sino porque carece de los mencionados incentivos sociales para tener tal comportamiento. En todo caso, hay que tener en cuenta que estas respuestas no las genera de manera espontánea, puesto que tiene un conocido y bien documentado sesgo a la complacencia.

Los economistas conocemos bien situaciones de esta naturaleza, en las que se tiene la capacidad para mejorar un resultado colectivo pero no se tienen los incentivos adecuados para hacerlo. En la teoría de juegos un *juego repetido* es aquel en el que los mismos participantes interactúan una y otra vez, de modo que lo que cada uno hace hoy condiciona el trato que los demás le dispensarán mañana. Es la situación de esta Academia, cuyos miembros coinciden cada semana durante años y son alternativamente ponentes y oyentes. En ese contexto, la cortesía recíproca induce a moderar las críticas y se sostiene por sí sola. Formular una objeción rigurosa puede conllevar un cierto coste individual, derivado de la incomodidad que pueda generar o del posible deterioro en las relaciones personales, a la vez que se asume el riesgo de recibir un trato equivalente cuando llegue el día de su exposición. En cambio, el beneficio de un debate más exigente que se podría conseguir con esta forma de proceder se distribuye al conjunto de los académicos, y no recae exclusivamente en quien ha asumido ese coste. Esta asimetría entre costes individuales y beneficios colectivos tiende a mantener un equilibrio de crítica tibia y moderada, que con mucha frecuencia muda en una sucesión de indiscriminados halagos que va mucho más allá de la corrección debida.

El turno de preguntas no solo puede dejar pasar estos problemas, sino que tiende a reproducirlos. Ocurre, por ejemplo, cuando una intervención, en lugar de dialogar con la ponencia, formula una pregunta ajena a la misma, o se convierte en una exposición erudita independiente. En este caso la erudición simplemente se exhibe, en vez de ponerse al servicio del juicio, y consume además un tiempo de la sesión que debiera corresponder al debate.

Lo que está en juego es la función misma de nuestras sesiones. Si cualquiera de nosotros puede interrogar directamente a un conjunto de fuentes y reconstruir un debate a su medida, la exposición oral de lo ya sabido pierde buena parte de su sentido, y confundir el formato de la ponencia-síntesis con la contribución que esta debería contener se convierte en una debilidad para la institución. Lo que justifica el hecho de reunirnos cada martes es, precisamente, lo que no puede obtenerse por esa vía, es decir, la tesis propia del ponente y la discusión que suscita. No me cabe duda alguna de que ese juicio lo tiene cualquier miembro de esta Academia, y la cuestión es si estamos dispuestos a expresarlo y a someterlo a debate.

Soy consciente de que una exhortación difícilmente bastará por sí sola para alterar aquel equilibrio al que me he referido, pero sí puede hacerlo una práctica que todos compartamos. Por ejemplo, con lo dicho hasta aquí, creo justificado sugerir que, en la medida de nuestras posibilidades, antes de exponer en el pleno el texto de nuestras ponencias lo sometamos a la crítica de un modelo de IA y respondamos a las objeciones que consideremos más fuertes.

La propuesta tiene, además, una ventaja que no es menor, ya que permite externalizar una parte de los costes personales de la crítica, sin renunciar a sus beneficios intelectuales. Al fin y al cabo, nadie puede sentirse excesivamente maltratado por haber recibido una objeción de un modelo de lenguaje. Naturalmente, como ya he señalado, la máquina no garantiza que sus críticas sean acertadas; corresponde al autor valorarlas y responder a las que considere fundadas.

A medio y largo plazo, todo ello podría modificar los criterios con que presentamos y debatimos nuestras ponencias, y también aquellos con los que valoramos el mérito académico. El mérito pasaría a depender menos de haber publicado un libro o presentado una ponencia y más de haber realizado una aportación original, ya sea teórica, empírica o interpretativa, cuyo valor se mantenga y sea reconocible, incluso cuando una IA pueda resumirla en pocos minutos.

A modo de epílogo

Termino volviendo a la coincidencia de títulos que mencioné al principio de la intervención. Alejandro Nieto señalaba en *El mundo visto a los noventa años* un rasgo compartido entre su obra y la de Ramón y Cajal: «su rechazo a cualquier tipo de erudición», pues ambos «cuentan simple y directamente lo que están viendo y pensando, sea trivial o profundo, y no lo que han leído de sabio alguno». Ramón y Cajal (1923) lo expresó con una imagen célebre al escribir: «rendimos tributo de veneración a quien añade una obra original a una biblioteca y se lo negamos a quien lleva una biblioteca en la cabeza».

Dando por supuesto que no pretendo establecer comparación alguna con ellos, comparto ese rechazo o, más exactamente, el rechazo a su mal uso; es decir, no a la erudición en sí, sino a la que sustituye al juicio propio. Citar no es el problema, el error estriba en no someter la información —a menudo manejada en exceso— a un filtro de coherencia y de relevancia para el propio argumento. Diferenciar estas dos formas de proceder es esencial y, además, muy factible hoy en día ya que contamos con herramientas que Ramón y Cajal no pudo concebir y que Nieto apenas llegó a conocer. Lo relevante no es la erudición acumulada, sino el saber interpretarla y ponerla a prueba con criterio. Esto cobra especial importancia en un tiempo en el que los modelos de lenguaje pueden reunir y ordenar casi al instante información y conocimiento que antes requerían años de estudio.

A los ochenta años, y desde la experiencia que he relatado, puedo reconocer en este proceso tanto su progresión gradual como su disrupción. Hoy contemplo la IA como la culminación de un deseo presente en mi larga trayectoria de relación con las máquinas, el de que aprendan de los datos en vez de tener que recibir reglas e instrucciones de nosotros. También es la mayor discontinuidad tecnológica de mi vida. Los modelos actuales persiguen ese objetivo con una cantidad de medios y a una escala que hace más de sesenta años eran, al menos para mí, inimaginables.

No sé hasta dónde llegará esta tecnología ni qué consecuencias alcanzará a tener, pero hoy me inspiran la misma reserva tanto el desmedido optimismo de unos como el excesivo pesimismo de otros. Sin embargo, y en nuestro ámbito académico, pienso que no es razonable la indiferencia de quienes descartan su utilización sin haberla ensayado alguna vez. Para mí es inevitable mirarla como quien vio empezar el proceso, y hacerlo a la vez con el asombro que me produce y con la prudencia que me han enseñado los años. No pretendo haber acertado en todo, pero en esta ocasión he intentado contar, como Ramón y Cajal o Nieto, lo que he visto y lo que pienso, y no lo que he leído de sabio alguno.

Referencias (enlaces validados en noviembre de 2026)

Chi, H., *et al*., «Unveiling Causal Reasoning in Large Language Models: Reality or Mirage?», *Advances in Neural Information Processing Systems*, vol. 37, 2024, pp. 96640-96670.

Knight, F. H., *Risk, Uncertainty and Profit*, Houghton Mifflin Company, Boston y Nueva York, 1921. https://fraser.stlouisfed.org/files/docs/publications/books/risk/riskuncertaintyprofit.pdf

Krizhevsky, A., Sutskever, I., y Hinton, G. E., «ImageNet Classification with Deep Convolutional Neural Networks», *Advances in Neural Information Processing Systems*, vol. 25, 2012, pp. 1097-1105.

Lighthill, J., *Artificial Intelligence: A General Survey*, Science Research Council, 1973.

McCarthy, J., Minsky, M. L., Rochester, N., y Shannon, C. E., «A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence», 1955.

Minsky, M. L., y Papert, S. A., *Perceptrons: An Introduction to Computational Geometry*, MIT Press, 1969.

Nieto García, A., *El mundo visto a los noventa años*, Editorial Comares, Granada, 2022.

Ramón y Cajal, S., «Cap. V: Enfermedades de la voluntad*», Reglas y consejos sobre investigación científica (Los tónicos de la voluntad),* discurso leído con ocasión de la recepción del autor en la Real Academia de Ciencias Exactas, Físicas y Naturales*,* 6.ª ed., Imprenta de Juan Pueyo, Madrid, 1923.
https://www.gutenberg.org/cache/epub/66373/pg66373-images.html

Ramón y Cajal, S., *El mundo visto a los ochenta años: impresiones de un arteriosclerótico*, Tipografía Artística, Madrid, 1934.

Rosenblatt, F., «The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain», *Psychological Review*, vol. 65, n.º 6, 1958, pp. 386-408.

Turing, A. M., «Computing Machinery and Intelligence», *Mind*, vol. 59, n.º 236, 1950, pp. 433-460.
https://phil415.pbworks.com/f/TuringComputing.pdf

Vaswani, A., *et al*., «Attention Is All You Need», *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998-6008.