

Do Good.
Do Better.

esade

2º semestre
2023 #33



El momento de la Inteligencia Artificial

Informe económico y financiero

Departamento de Economía, Finanzas y Contabilidad

esade

EsadeEcPol - Center
for Economic Policy

CON EL APOYO DE

 **Sabadell**

La inteligencia artificial desde la perspectiva de la economía del comportamiento

Anna Bayona

Profesora titular del Departamento de Economía, Finanzas y Contabilidad de Esade y Directora del Grupo de Investigación en Economía y Finanzas (GREF)

En 1950, el matemático inglés Alan Turing se preguntaba en un artículo: “¿pueden las máquinas pensar?”. Variando un tanto esa formulación, también se planteaba “si es posible que las máquinas muestren un comportamiento inteligente”.¹ Desde entonces, estas preguntas han revolucionado la historia de la ciencia y constituido el germen del desarrollo de la Inteligencia Artificial (IA). En el tiempo transcurrido, la IA se ha desarrollado enormemente, hasta conseguir que los ordenadores aprendan, razonen, utilicen el lenguaje, perciban y resuelvan problemas, funciones que suelen asociarse a la inteligencia humana. La cantidad de datos y la capacidad de procesamiento utilizadas para entrenar los sistemas de IA se han multiplicado por cien millones en los últimos diez años.² Esta gran cantidad de datos, junto con ordenadores más rápidos y nuevos avances en los algoritmos, han permitido que el desarrollo de la IA alcance un nuevo nivel: sus usos, que ya tienen un efecto significativo en la economía, la mayoría de los sectores, la sociedad y la humanidad, van desde los vehículos autónomos hasta aplicaciones que pueden leer, entender y transmitir lenguaje, como los *chatbots*.

¹ Turing (1950).

² Hogarth (2023).

Este artículo analiza cómo la interacción entre el ser humano y la IA afecta a la organización y el comportamiento económicos. Se trata de una cuestión con muchas facetas, pero aquí adoptaremos la perspectiva de la economía del comportamiento. Es decir, nos centraremos en el conocimiento actual sobre cómo los individuos toman decisiones desde el punto de vista de la teoría económica, que se ve enriquecido por disciplinas como la psicología, la sociología y los estudios culturales, entre otras. La economía conductual suele utilizar experimentos para apoyar o refutar hipótesis sobre las creencias y las elecciones de las personas.

La cantidad de datos y la capacidad de procesamiento utilizadas para entrenar los sistemas de IA se han multiplicado por cien millones en los últimos diez años.

La inteligencia artificial y el comportamiento humano

Una buena manera de empezar el análisis de la interacción entre los humanos y las máquinas en contextos económicos es centrarse en la manera en que se comportan en un experimento las personas cuando interactúan con agentes informáticos.³ Existen pruebas concluyentes de que la interacción entre actores humanos y agentes informáticos suele cambiar el comportamiento estratégico de los primeros. En concreto, los sujetos humanos se adaptan a los agentes informáticos, incluso cuando no se les proporciona información sobre ellos. Además, cuando interactúan con agentes informáticos se comportan más racionalmente y, en general, de manera más egoísta. Esto plantea la cuestión de cómo cooperan los humanos con la IA frente a cómo lo hacen con otros humanos. Hay pruebas de que los humanos cooperan menos con actores de IA benévolo que con actores humanos, y que esto solo ocurre si satisface sus intereses egoístas. El aprendizaje que podemos obtener de esto es que, cuando los humanos interactúan con ordenadores, tanto el comportamiento individual como los resultados económicos suelen ser diferentes. Lo cual suscita preguntas sobre cómo evolucionarán la economía y la sociedad cuando muchas interacciones son entre humanos e IA, o entre IA e IA.

Varios estudios demuestran que, cuando la interacción se produce entre humanos, la gente tiene preferencias sociales; es decir, que los individuos no solo actúan en interés propio, como asumen los modelos económicos tradicionales, sino que tienen preferencias positivas o negativas por los beneficios de otras personas.⁴ En concreto, los individuos prefieren la justicia recíproca, lo que significa que nuestras acciones pasadas influyen en cómo nos tratan los demás. Además, la identidad de una persona (es decir, la conciencia que tiene de sí misma y la percepción que los demás tienen de ella) influye en las elecciones individuales, las interacciones y los resultados económicos.⁵ Por ejemplo, se ha demostrado que la identidad y la costumbre explican las contribuciones a la beneficencia y las donaciones de antiguos alumnos, las contribuciones a bienes públicos, la publicidad y la elección de colegio, entre otras cosas. En la medida en que los ordenadores y la IA influyen en cómo

³ March (2021).

⁴ El influyente trabajo de Fehr y Fischbacher (2002).

⁵ Akerlof y Kranton (2000).

Cuando los humanos interactúan con ordenadores, tanto el comportamiento individual como los resultados económicos suelen ser diferentes.

experimentamos el mundo, sigue siendo una cuestión no resuelta la manera en que la IA afectará al concepto que tenemos de nosotros mismos y a aquellas características que son intrínsecamente humanas, como por ejemplo la conciencia, las emociones, las intuiciones, la creatividad, las conversaciones, el libre albedrío y el razonamiento moral.

Además, como han documentado economistas del comportamiento como Kahneman, Tversky, Thaler o Shiller, los individuos suelen tomar decisiones que no son plenamente racionales. Las personas utilizan heurísticos para tomar decisiones y tienen, frente al *Homo economicus*, sesgos como la contabilidad mental, el efecto de anclaje, el encuadre, el comportamiento gregario, el sesgo de presente, el sesgo de actualidad, el sesgo de confirmación, el sesgo de familiaridad, el sesgo de *statu quo* y la falta de atención, entre muchos otros. Algunos de ellos han demostrado ser potencialmente perjudiciales para el bienestar personal y económico de las personas.⁶ Se podría pensar, pues, que tal vez la IA permita eliminar todos los posibles sesgos y heurísticos y hacer realidad el *Homo economicus*.

Sin embargo, cabe preguntarse si estos heurísticos y sesgos han resultado útiles en nuestra senda evolutiva hacia el *Homo sapiens*. En una investigación muy interesante, Chen, Lakshminarayanan y Santos (2006) estudian el comportamiento de compra de los monos capuchinos cuando se les pide que distribuyan un presupuesto de fichas entre distintos tipos de alimentos. Descubren que los monos capuchinos actúan racionalmente ante las alteraciones de

⁶ Sin embargo, los sesgos de comportamiento no siempre son perjudiciales para la utilidad y el bienestar de los individuos. En contextos de mercado relacionados con la electricidad y la compraventa en los mercados financieros, Bayona y Manzano (2022) muestran que, en ciertas condiciones, un tipo de sesgo de comportamiento puede aumentar los beneficios y el bienestar de los vendedores.

precio y de riqueza, pero tienen sesgos como la dependencia de referencia y la aversión a la pérdida. Esto indica que hay sesgos del comportamiento que no son únicamente humanos y son innatos; es decir, no aprendidos. De hecho, algunos académicos han demostrado que algunas de estas preferencias aparentemente no racionales tienen cierto valor evolutivo.⁷ Por lo tanto, parece peligroso dejar que la IA elimine la mayoría de los sesgos de comportamiento del ser humano en la toma de decisiones. De cara al futuro, cabe plantearse qué efectos tendrá en nuestro cerebro el hecho de delegar muchas tareas intelectuales en la IA. El escritor estadounidense Nicholas Carr sostiene que el uso de internet está reduciendo nuestra capacidad de leer y pensar de manera crítica, y que la IA puede afectar a nuestra percepción de la realización personal y nuestra felicidad. Es posible imaginar un futuro en el que los humanos tengan que ir al gimnasio intelectual para ejercitar sus cerebros no entrenados.

Herbert Simon, pionero del estudio multidisciplinar de la IA y las ciencias del comportamiento, fue el primero en abordar simultáneamente la informática y la economía. De hecho, recibió premios en ambos campos: el Premio Turing de informática en 1975 y el Premio Nobel de Economía en 1978. El trabajo de Simon señalaba que la racionalidad humana tiene límites; una idea influida por su estudio de la IA. Creía que el análisis de las simulaciones por ordenador era una manera útil de estudiar la cognición humana. Este enfoque también puede aplicarse al contexto actual: ahora, la IA puede contribuir a la investigación de la economía del comportamiento intentando buscar nuevos modelos de patrones de comportamiento que afectan a las elecciones humanas.

⁷ Samuelson y Swinkels (2006).

De cara al futuro, cabe plantearse qué efectos tendrá en nuestro cerebro el hecho de delegar muchas tareas intelectuales en la IA.

La inteligencia artificial como creadora y tomadora de decisiones

La IA tiene una enorme influencia en nuestras vidas, porque ya toma decisiones automatizadas con escasa intervención humana. Estas van desde los procesos de producción hasta los servicios; desde la decisión de conceder un préstamo hasta la selección de un candidato para un puesto de trabajo. Además, los sistemas de IA generativa más recientes pueden crear textos, imágenes, música, vídeos o código y, por lo tanto, ampliar los tipos de tareas que puede hacer para la gente con trabajos creativos. La IA sigue desarrollándose para abarcar más aspectos de la inteligencia artificial general (IAG), que, en el futuro, será capaz de hacer la mayoría de las tareas que hacen los humanos. Los beneficios de la IA para la economía son potencialmente enormes, puesto que puede ahorrar costes; sus procesos están automatizados y son más rápidos, las decisiones son más coherentes que si las tomaran personas y puede ayudar a los humanos con ciertos sesgos y heurísticos (por ejemplo, la atención). Más importante aún, la IA puede ayudarnos a abordar algunos de los retos a los que se enfrenta la humanidad, como la crisis climática o la curación de determinadas enfermedades. Por lo tanto, si las herramientas de IA se gestionan de manera adecuada, tienen el potencial de transformar positivamente nuestra economía y nuestras sociedades.

La IA sigue desarrollándose para abarcar más aspectos de la inteligencia artificial general (IAG), que, en el futuro, será capaz de hacer la mayoría de las tareas que hacen los humanos.

Sin embargo, muchas veces las decisiones automatizadas que toman los algoritmos de IA difieren de las decisiones humanas. Eso lleva a la pregunta de si las decisiones algorítmicas de la IA pueden ser justas (obsérvese que no sostengo que las humanas siempre lo sean). Los académicos suelen distinguir entre objetivos algorítmicos sesgados y predicciones algorítmicas sesgadas.⁸ Los algoritmos tienen normalmente una función objetivo (como, por ejemplo, maximizar los beneficios), pero esta puede estar sesgada, porque implica *trade-offs* que requieren cierta noción de justicia. Por supuesto, esto abre un largo debate filosófico sobre qué preferencias deben representar estos algoritmos. Lo cual es muy relevante, porque buena parte de los algoritmos de IA pertenecen a unas pocas empresas cuyos intereses privados pueden no coincidir con los de la sociedad. En cuanto a las predicciones algorítmicas, también podrían estar sesgadas porque, por ejemplo, hay infrarrepresentaciones en las muestras de datos que se utilizan para entrenar a los algoritmos (si solo se utilizaran datos del desempeño de los candidatos que fueron contratados, pero no de los que no lo fueron). Si los datos de entrenamiento están mal etiquetados, esto podría generar aún más sesgos. Por ejemplo, un trabajador al que su jefe discrimina podría ser etiquetado como de bajo rendimiento, con lo que la predicción del algoritmo generará una discriminación todavía mayor. También podría suceder que los propios programadores informáticos que escriben el código tengan sesgos y que se produjeran bucles de retroalimentación algorítmica, lo que significa que las predicciones afectan causalmente al resultado previsto. Muchas de las predicciones algorítmicas que están sesgadas tienen soluciones técnicas que probablemente se resolverán con el desarrollo de la IA. Sin embargo, los objetivos algorítmicos sesgados no tienen una solución técnica y requieren regulación y el consenso del conjunto de la sociedad. Esto es relevante, porque los humanos estamos delegando decisiones importantes en las máquinas.

La IA también plantea algunos retos que tienen que ver con la utilización de datos personales y los derechos de propiedad. Muchos sistemas de IA requieren técnicas que procesan datos personales para predecir el comportamiento y los logros de una persona (elaboración de perfiles). Eso puede suponer muchos problemas para la privacidad de nuestros datos y nuestras vidas, porque se utiliza y queda registrada cualquier actividad digital. Además, la IA generativa implica dos tipos de riesgos. El primero está relacionado con los derechos de propiedad intelectual. Determinar a quién

⁸ Cogwill y Tucker (2020).

Los objetivos algorítmicos sesgados no tienen una solución técnica y requieren regulación y el consenso del conjunto de la sociedad.

pertenece el contenido creado por la IA y si este contenido puede infringir las leyes de derechos de autor, patentes y marcas, sigue siendo una cuestión sin resolver. El segundo es un problema más básico de los modelos de IA generativa que aplican el aprendizaje automático de un conjunto de datos. Estos pueden proporcionar respuestas muy precisas capaces de obtener una buena nota en un examen de nivel universitario y, al mismo tiempo, inventar hechos y relatos que son falsos. ¿Y si la IA se usa para generar noticias falsas que beneficien a una parte de la sociedad? ¿Y si a la sociedad le resulta cada vez más difícil distinguir la verdad de la falsedad? ¿Y si no nos damos cuenta de que estamos hablando con un interlocutor de IA? Yuval Noah Harari sostiene que la IA podría utilizarse para destruir democracias porque mermará la capacidad de conversar entre nosotros.⁹

Entender cómo crea contenidos o toma decisiones la IA implica comprender el proceso mediante el cual funciona. Esto es importante, porque gran parte de la ciencia se dedica a entender la causa y el efecto y el proceso a través del cual esto ocurre. Sin embargo, hay tipos de IA que no comprenden siquiera sus programadores.¹⁰ ¿Podemos abrir la caja negra de la IA? Del mismo modo que los neurocientíficos estudian el cerebro, los informáticos investigan ahora formas de entender cómo toma decisiones la IA. Pero si a los informáticos les resulta difícil, ¿cómo esperar que la sociedad en general entienda las decisiones o las recomendaciones de la IA? Existen pruebas concluyentes de que los humanos somos aversos a la ambigüedad; es decir, preferimos los riesgos conocidos a los desconocidos. Lo cual plantea la cuestión de si nuestra sociedad aceptará que decisiones importantes sean tomadas por algoritmos de IA.

⁹ Harari (2023).

¹⁰ Castelvecchi (2016).

La inteligencia artificial: popularidad y adopción

En 2022, el Centro de Investigación Pew entrevistó a los estadounidenses sobre su opinión acerca de la IA.¹¹ La encuesta preguntaba cómo les hace sentir el uso cada vez mayor en la vida cotidiana de programas informáticos de IA: el 37% de los adultos estadounidenses respondieron que están más preocupados que entusiasmados; el 45% que preocupados y entusiasmados en la misma medida; y el 18% que más entusiasmados que preocupados. Quienes están más preocupados que entusiasmados mencionan como razones la pérdida de puestos de trabajo para los humanos; la vigilancia, los hackeos y la privacidad digital; y la falta de conexión y cualidades humanas. Los que están más entusiasmados que preocupados aducen que la IA mejora la vida y la sociedad; ahorra tiempo y es más eficiente; y que su progreso inevitable es el futuro. Un análisis más detallado revela que a los estadounidenses les preocupan sobre todo los sistemas de IA que conocen los pensamientos y el comportamiento de las personas, y los que toman importantes decisiones vitales en nombre de la gente. Los resultados de la encuesta no muestran una opinión uniforme respecto a la IA, aunque la mayoría de los estadounidenses tienden más a la preocupación que al entusiasmo.

Estas opiniones nos hacen preguntarnos cómo se utiliza actualmente la IA en las organizaciones. La encuesta global de McKinsey sobre IA de 2022 mostró que el 50% de las organizaciones que componen su variada muestra ya habían adoptado la IA. Las funciones de IA más utilizadas eran: la automatización de procesos robotizados, la visión artificial, la comprensión de textos en lenguaje natural y los agentes virtuales o las interfaces conversacionales.¹² También reveló que, entre las organizaciones que utilizan la IA, los casos de uso más importantes se dan en la optimización del funcionamiento de servicios, la creación de nuevos productos basados en IA, la mejora de productos basada en IA, el análisis de la atención al cliente y la segmentación de clientes. Como puede verse, muchas de estas funciones son propias de los departamentos de operaciones y marketing de las empresas. Sin embargo, del informe se deduce algo llamativo: que las empresas no han mejorado sus estrategias para reducir los riesgos de la IA, por ejemplo, la ciberseguridad, el cumplimiento normativo, la privacidad personal, la explicabilidad

¹¹ Rainie, *et al.* (2022).

¹² McKinsey & Company (2022).

La respuesta normativa de los distintos gobiernos y organismos internacionales será fundamental para equilibrar los beneficios potencialmente extraordinarios de la IA y sus notables riesgos.

de los modelos de IA, la reputación de la organización, la equidad y la justicia, el desplazamiento laboral y la seguridad física. Esto cuestiona la existencia de incentivos privados para reducir los riesgos que plantea la IA, y evidencia que la divergencia entre los incentivos privados y los sociales exige la regulación de la IA.¹³ La respuesta normativa de los distintos gobiernos y organismos internacionales será fundamental para equilibrar los beneficios potencialmente extraordinarios de la IA y sus notables riesgos.

El futuro

Muchos especulan sobre el futuro de la IA y sus efectos en la sociedad, la economía y la humanidad. Este artículo se ha centrado en el impacto que tiene la IA en el cerebro humano, la toma de decisiones y la interacción con los demás. En conclusión, si se utiliza y regula adecuadamente, la IA puede ayudar a la humanidad a prosperar, al contribuir a resolver los complejos problemas de nuestro mundo. Sin embargo, los humanos no debemos dejar de ejercitar nuestros cerebros y tenemos que seguir leyendo, escribiendo, contando, además de practicar el pensamiento crítico, la creatividad, comunicarnos y fomentar el entendimiento. Solo así cultivaremos nuestra naturaleza humana intrínseca •

¹³ Varios pensadores importantes han advertido sobre la posibilidad de que la IA plantee riesgos existenciales para la humanidad, entre ellos Stephen Hawking, Yuval Noah Harari, George Soros, Bill Gates o Elon Musk.

Referencias

- Akerlof, G. A., y Kranton, R. E. (2000). "Economics and Identity". *The Quarterly Journal of Economics*, 115(3), 715-753.
- Bayona, A., y Manzano, C. (2022). "Competition in Schedules with Traders that Neglect the Informational Content of the Price." Disponible en SSRN: <https://ssrn.com/abstract=4085637> o <http://dx.doi.org/10.2139/ssrn.4085637>.
- Carr, N. (2014). *The Glass Cage: Automation and Us*. W. W. Norton & Co.
- Castelvecchi, D. (2016). "Can we open the black box of AI?" *Nature News*, 538(7623), 20.
- Chen, M. K., Lakshminarayanan, V., y Santos, L. R. (2006). "How basic are behavioral biases? Evidence from capuchin monkey trading behavior". *Journal of Political Economy*, 114(3), 517-537.
- Cowgill, B., y Tucker, C. E. (2020). "Algorithmic fairness and economics". *Columbia Business School Research Paper*.
- Fehr, E., y Fischbacher, U. (2002). "Why social preferences matter – the impact of non-selfish motives on competition, cooperation and incentives". *The Economic Journal*, 112(478), C1-C33.
- Harari, Y. N. (28 de abril de 2023). "Yuval Noah Harari argues that AI has hacked the operating system of human civilisation". *The Economist*.
- Hogarth, I. (13 de abril de 2023). "We Must Slow Down the Race to God-Like AI". *FT Magazine*. <https://www.ft.com/content/03895dc4-a3b7-481e-95cc-336a524f2ac2>.
- March, C. (2021). "Strategic interactions between humans and artificial intelligence: Lessons from experiments with computer players". *Journal of Economic Psychology*, 87, 102426.
- McKinsey & Company (2022). The state of AI in 2022 – and half a decade in review. QuantumBlack by McKinsey & Company.
- Samuelson, L., y Swinkels, J. M. (2006). "Information, evolution and utility". *Theoretical Economics*, 1(1), 119-142.
- Rainie, L., Funk, C., Anderson, M., y Tyson, A. (17 de marzo de 2022). "How Americans Think about Artificial Intelligence". Centro de Investigación Pew. <https://www.pewresearch.org/internet/2022/03/17/how-americans-think-about-artificial-intelligence/>.
- Turing, A. M. (1950). "Computing Machinery and Intelligence". *Mind*, 59(236), 433-460.

esade

esade

EsadeEcPol - Center
for Economic Policy

Informe económico y financiero

Departamento de Economía, Finanzas y Contabilidad

CON EL APOYO DE

 **Sabadell**