

El toque humano. ¿Cómo decidirá un coche autónomo entre la supervivencia del conductor o de un peatón? ¿O un algoritmo sobre la concesión de una hipoteca? La tecnología reabre el debate sobre el 'interés general'

LA ÉTICA DEL ROBOT

CÓMO LOGRAR
QUE LAS
MÁQUINAS
TENGAN 'ALMA'

POR PABLO PARDO
ILUSTRACIÓN: AKIRANT

Suponga que usted está conduciendo un tranvía. Da una curva y ve delante a cinco obreros que están trabajando en la vía. Usted aprieta el freno pero el tranvía no se detiene. Entonces ve un desvío: usted puede encaminar el tranvía en esa dirección y salvar a los cinco hombres. Pero, desgraciadamente, en esa otra sección hay un obrero. Si usted mete al tranvía allí, matará a esa persona, pero salvará a las otras cinco.

Ahora suponga que usted es un cirujano de trasplantes. Hay cinco enfermos que morirán si no reciben órganos hoy. Dos precisan un pulmón cada uno; otros dos, sendos riñones; el quinto, un corazón. Entonces llega a la clínica, para un chequeo rutinario, un hombre joven y sano, cuyo tipo sanguíneo coincide con el de los cinco enfermos. Es un donante ideal. Lo único que tiene que hacer es matarlo y distribuir sus órganos entre los que lo necesitan.

¿Qué haría usted? Estos dilemas fueron popularizados en un artículo publicado en 1985 por Judith Jarvis Johnson en la revista *The Yale Law*

Journal, de la Facultad de Derecho de Yale, aunque en realidad fueron formulados por primera vez en 1905. Pero la respuesta no ha variado, casi de forma unánime, todos vemos ético matar a la persona en la vía del tren, pero no a la del hospital. Es una reacción casi automática, pese a que en términos puramente cuantitativos se trata de lo mismo: acabar con una vida para salvar cinco.

La razón es que en el primer caso no hay contacto físico; en el segundo, sí. El

tranvía —una máquina— acabará con la vida de un ser humano al que no conocemos, y eso es diferente a matar con nuestras propias manos.

Así que la frase «una muerte es una tragedia; un millón de muertes, una estadística», atribuida (erróneamente) a Stalin es cierta. La idea de matar nos

repugna (a la mayoría). Pero más aún la idea de matar personalmente.

Ahora bien, ¿cómo se puede programar a una máquina para que ésta tome decisiones similares?

No es una pregunta teórica. La Inteligencia

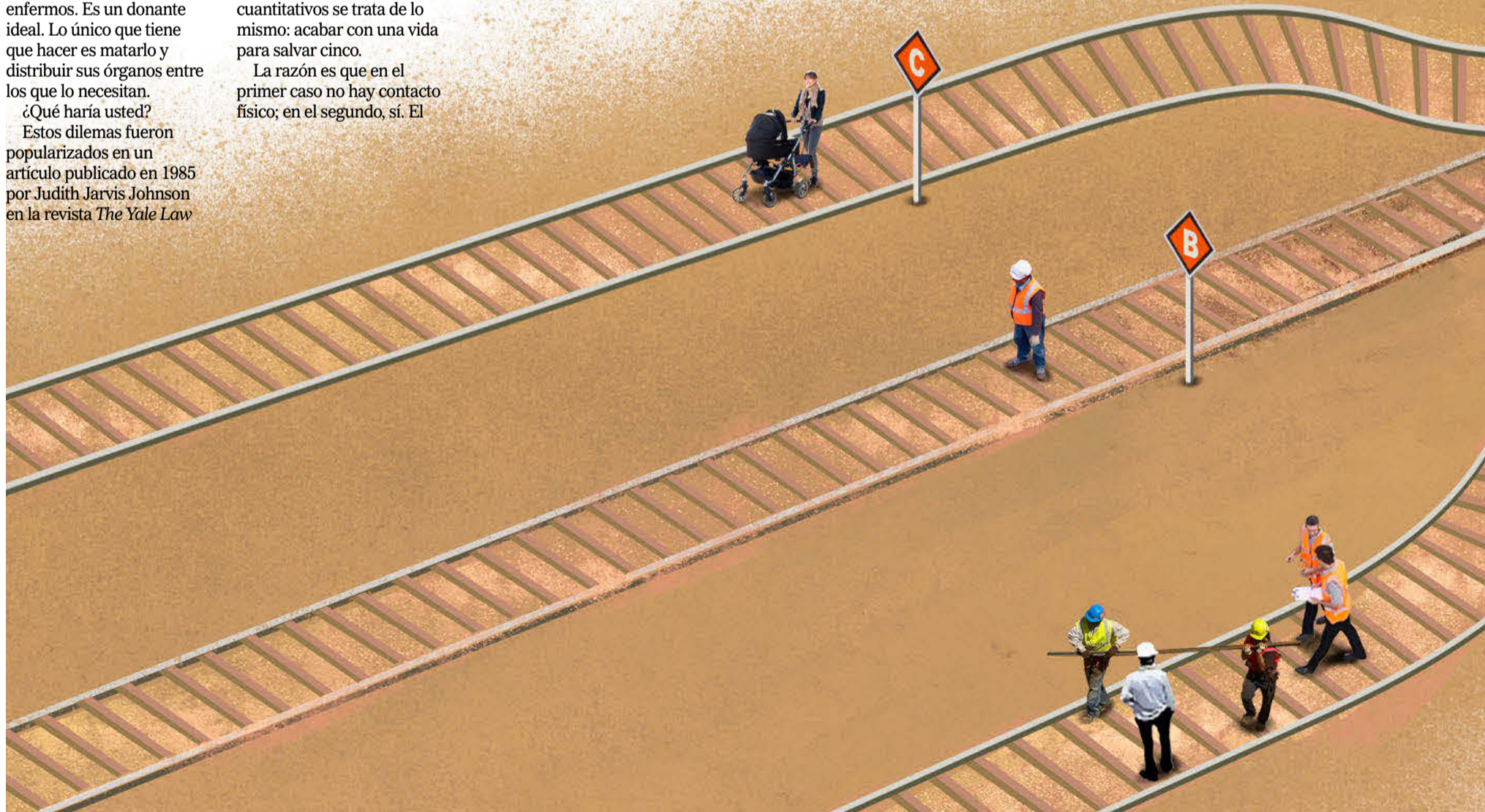
Artificial (IA) y sus versiones previas, los algoritmos de los programas de ordenador, ya están tomando decisiones que afectan a nuestra vida. Un programa informático ya decide si usted recibirá un crédito o no. También puede decidir si usted va a ser desahuciado o no: el viernes de la semana pasada, Wells Fargo, el tercer mayor banco de EEUU admitió que, por un error informático, 600 clientes habían visto denegada la renegociación de su hipoteca. A consecuencia de ello, 400 de ellos fueron desahuciados de sus casas.

Un programa informático también puede decidir dónde va a ir el dinero de su plan de pensiones. La gestora Vanguard tiene 112.000 millones de dólares (97.000 millones de euros)

gestionados por *roboadvisors*, es decir, *roboasesores*: robots que deciden dónde se invierte el dinero. Desde 2010, el índice S&P500 de las grandes empresas de Wall Street ha subido un 145%, pero el número de asesores financieros ha caído en un 3%. ¿La razón de la caída? Los robots.

Y ¿dónde empieza la ética de los robots? Por volver al ejemplo inicial, ¿qué pasa si el tranvía no tiene conductor, como los Metros de Vancouver o Singapur o, en el futuro, los coches autónomos?

Este último caso es ya un factor de incertidumbre para una industria que mueve cada año 700.000 millones de euros: la de los seguros de automóvil. ¿Cómo se puede evaluar la



responsabilidad de un software?

Ni siquiera hace pensar en casos extremos en los que el coche deba decidir, por ejemplo, si mata al conductor o a un peatón. Imaginemos un mundo de coches autónomos en el que cada vehículo tiene su seguro correspondiente. Ahora, supongamos que hay usuarios que tienen mejores seguros y que, además, hacen que sus coches informen a los otros vehículos de que, si chocan con ellos, les va a salir más caro a sus dueños. En principio, los demás vehículos tratarán de evitar colisiones con esos

automóviles. Así, la Inteligencia Artificial (IA) habrá creado conductores de primera y de segunda.

En la IA, la persona programa a la máquina. Por volver a los ejemplos iniciales, no tiene que decidir si mata o no personalmente. La ética puede quedar reducida a un problema abstracto, salvo para el que muere, claro está, como deja de manifiesto la web Moralmachine (*Máquinamoral*) del Instituto Tecnológico de Massachusetts (MIT, por sus

en el caso de expertos, la complejidad de los algoritmos y de los sistemas de toma de decisiones automática es tal que sólo algunos expertos pueden comprenderlo», explica Pablo Molina, profesor de Ética y Tecnología de la Universidad de Georgetown. Como recalca Molina, «al crear los programas, existe el riesgo de que las personas trasladen sus prejuicios a las máquinas».

demonstró que COMPAS daba una mayor puntuación y, por tanto, más posibilidad de reincidencia a las personas de raza negra que a las de raza blanca».

Un sistema más extremo está a punto de ser desplegado en la segunda potencia mundial. Dentro de dos años, China tendrá 600 millones de cámaras para monitorizar a sus 1.400 millones de ciudadanos. Y la empresa Cloud Walk está ensayando un sistema de seguimiento de las personas con esas cámaras para determinar si éstas van a cometer un delito. «Si alguien compra un cuchillo de cocina, vale. Pero si esa persona compra después un saco y un martillo, se convierte en sospechosa»,

quedado de manifiesto con Palantir, una empresa propiedad del cofundador de eBay y miembro del consejo de administración de Facebook, Peter Thiel. En abril, la revista *Bloomberg Business Week* publicó un reportaje sobre Palantir que incluía el análisis de uno de sus productos, Gotham, diseñado por la consultora Justice & Security Strategies y empleado por la Policía de Los Angeles. Gotham usa una serie de datos para evaluar si los ciudadanos a los que la policía pone en su base de datos pueden ser delincuentes o no. Los

ser parte de ningún grupo delictivo. Aun así, Palantir ha decidido que es sospechoso por asociación.

Más surrealista es el experimento que llevó a cabo en julio Jacob Snow, del grupo de izquierdas



siglas en inglés), que plantea, en un formato de juegos, una serie de problemas éticos de las nuevas tecnologías.

Además, la IA, como casi todas las nuevas tecnologías, no está regulada. «Hemos creado un monstruo de Frankenstein porque, salvo

**EL BANCO WELLS
FARGO ADMITIÓ
QUE UN ERROR
INFORMÁTICO
PROVOCÓ QUE 400
CLIENTES FUERAN
DESHAUCIADOS**

Hace 14 meses, el Tribunal Supremo del estado de Wisconsin, en EEUU, rechazó admitir a trámite el caso de Eric L. Loomis, que había sido sentenciado a seis años de cárcel por robar un coche. El juez que le condenó fijó esa condena usando el software COMPAS, cuyo algoritmo es secreto, y que ha sido desarrollado por la empresa Northpointe. La clave del COMPAS es que mide las posibilidades de que una persona vaya a reincidir. Es decir: Loomis fue condenado en parte por lo que la inteligencia artificial dijo que podría hacer en el futuro.

COMPAS, además, refleja los sesgos de sus creadores. Como explica el abogado de Cremades&Calvo-Sotelo Diego Solana, «posteriormente, a través de técnicas de ingeniería inversa, otro estudio

declaró en 2017 un portavoz de Cloud Walk al diario *Financial Times*. Cuando el nivel de sospecha alcanza un cierto nivel, Cloud Walk avisa a la policía.

Pekín podrá de ese modo encarcelar no sólo a quien se le oponga, sino a quien podría oponerse en el futuro. Así lo dijo hace un año el viceministro chino de Ciencia y Tecnología, Li Meng: «Si empleamos nuestros sistemas inteligentes bien, podremos saber de manera anticipada quién va a hacer algo malo».

Así pues, el programador traslada sus sesgos éticos a la máquina. Pero ésta también se ve influida por otras variables. «La clave para determinar los resultados que a nuestras preguntas nos dé el sistema de inteligencia artificial estará totalmente condicionado por los datos sobre los que la máquina haya trabajado», explica Solana.

Eso plantea más problemas, como ha

datos incluyen el historial delictivo, pero, también, lo que Sarah Brayne, de la Universidad de Texas, llama «supervisión secundaria» (*secondary surveillance*), como la actividad en redes sociales de la persona, su historial crediticio, los peajes que ha pagado en autopistas, sus compras por internet o por teléfono (incluyendo pizzas) y hasta los vídeos de cámaras de lugares como hospitales o parkings.

Así es como, según esta revista, Manuel Rios, de 22 años, ha sido detenido más de una docena de veces en los últimos dos años después de que un policía tomara una foto de él con un amigo que pertenecía a la banda callejera 18th Street Gang y la enviara a Gotham. Rios no parece

**UNA EMPRESA
USA DATOS DE
CONSUMO PARA
EVALUAR QUIENES
PUEDEN ACABAR
COMETIENDO
UN DELITO**

Unión por las Libertades Civiles en EEUU (ACLU, según sus siglas en inglés), cuando aplicó el programa Rekognition, que el gigante de internet Amazon está dando gratis a las policías de Florida y Oregón, a fotos de todos los miembros del Congreso de EEUU. Rekognition identificó como «delincuentes» a 28 congresistas. La mayoría, hispanos y negros.

Finalmente, está el riesgo de que estos sesgos refuercen los de las personas que reciben los datos procesados por sistemas de inteligencia artificial. Es lo que se llama «sesgo por automatización» (*automation bias*). En otras palabras: alguien puede ser considerado sospechoso por quien ha hecho un programa, el programa puede emplear los datos disponibles para reforzar esa sospecha y, finalmente, por desidia, la persona que evalúe la información del programa puede aceptar sus conclusiones sin más.

Volviendo a los ejemplos iniciales, es una cascada de decisiones abstractas pero con consecuencias muy reales. Los casos de COMPAS, Gotham, Cloud Walk, y Rekognition muestran que la inteligencia artificial necesita urgentemente recibir unas clases de ética.