Ética y algoritmos: una combinación necesaria

Vivimos rodeados de algoritmos que nos aconsejan qué comprar, nos ayudan a llegar a nuestro destino, fijan un precio para nuestros viajes o indican si somos merecedores de un crédito. ¿Podemos estar seguros de que esos algoritmos son neutrales y no responden a intereses ocultos?

Desde finales de 2017, los titulares de los periódicos cada vez parecen más sacados de una novela de ciencia ficción: desinformación y noticias falsas, criptomonedas, inteligencia artificial, bots, redes sociales... y, por debajo, una palabra recurrente: algoritmos. Simplificando mucho, un algoritmo no es más que una secuencia de pasos definidos que, al seguirla, nos permite resolver un problema determinado. Por ejemplo, calcular una raíz cuadrada a mano, cocinar una fabada o llegar hasta el parque más cercano. Basta con seguir las instrucciones en orden para obtener el resultado, ya sea un número, un plato de comida o llegar a nuestro destino.

Cada vez más vivimos inmersos en la cultura del algoritmo aunque no seamos conscientes de ello. Cuando le pedimos a nuestro ordenador o a nuestro teléfono que nos indique dónde hay una peluquería abierta a estas horas o que nos recomiende un restaurante cercano, estamos poniendo en marcha un algoritmo que recoge datos (dónde estamos, qué hora es, qué otros restaurantes hemos visitado, ...) los transmite a un servidor central donde los combina con los datos de millones de usuarios, ejecuta una serie de operaciones matemáticas y nos devuelve la lista de restaurantes que nos pueden gustar. Incluso nos puede generar un algoritmo personalizado (la ruta que tenemos que seguir en coche o en transporte público para llegar a ese restaurante).

Sin percatarnos de ello, el contexto técnico ha empezado a permear en el contexto social. Lo que antes eran ámbitos separados, tecnología y sociedad, que coincidían de manera tangencial, ahora cada vez están más relacionados entre sí y empieza a ser imposible separar uno del otro. Al hablar de la tecnología más cercana hemos pasado de hablar de un nuevo dispositivo, un virus o un "fallo informático" a hablar de conceptos más abstractos: los coches autónomos y la responsabilidad ante un accidente, el reconocimiento de rostros como herramienta de control y cómo podríamos engañar a una máquina para burlar la seguridad o la vigilancia. En estas conversaciones aparece por primera vez un factor que se había ignorado: la ética, en este caso, aplicada a los algoritmos y, por extensión, a los ordenadores que toman o nos ayudan a tomar las decisiones. Pasamos de estudiar lo correcto o equivocado del comportamiento humano a estudiar lo correcto o equivocado del comportamiento (de las decisiones) que toman "las máquinas", ése ente abstracto y amenazador que, en el fondo, no son más que ordenadores programados por seres humanos.

> Las páginas que visitamos, los enlaces que pulsamos y los que no, ... reflejan nuestros deseos, nuestros

movimientos, y acaban siendo analizadas por algoritmos para predecir nuestro potencial como trabajadores, como amantes, como estudiantes o como criminales

Nos hemos convertido en unos generadores de datos. Cualquier acción (o inacción) que hagamos *online* queda registrada, y será evaluada junto con las acciones de millones de usuarios, para categorizarnos, agruparnos, clasificarnos y proporcionar el bien o el servicio que mejor se adapta a nuestro perfil. Las páginas que visitamos, los enlaces que pulsamos y los que no, el tiempo que pasamos mirando una foto, si deslizamos hacia la derecha o hacia la izquierda… todas esas microinteracciones reflejan nuestros deseos, nuestros movimientos, nuestras dudas, y acaban siendo analizadas por algoritmos para predecir nuestro potencial como trabajadores, como amantes, como estudiantes o como criminales.

Un ordenador puede evaluar en milésimas de segundo una cantidad ingente de solicitudes de hipotecas o de currículos, y ofrecer las mejores condiciones (para el banco) o decidir si un candidato es el más adecuado. Desde el punto de vista empresarial es muy tentador: se pulsa una tecla y el sistema nos devuelve una lista de candidatos, ordenados del más prometedor al menos. Es algo limpio, aséptico, sin la posibilidad de que el departamento correspondiente, formado por humanos falibles, pueda incorporar sus prejuicios. El resultado es matemático, objetivo, imparcial. Lo dice la máquina-oráculo y no podemos dudar de su sabiduría ni de su asepsia.

Los algoritmos pueden ser tan subjetivos como los científicos que los han programado, y reflejan, de manera consciente o no, los sesgos y los prejuicios de los autores

Sin embargo esta percepción es falsa. Los algoritmos pueden ser tan subjetivos como los científicos que los han programado, y reflejan, de manera consciente o no, los sesgos y los prejuicios de los autores, que elaboran el modelo con la mejor de las intenciones. Este sesgo se puede dar en diferentes puntos del proceso.

El primer punto donde puede aparecer la parcialidad es al definir los datos que elegimos recoger o las preguntas que hacemos para obtener esos datos. El conjunto de datos elegido o la forma en que formulamos las preguntas puede no ser imparcial y guiar al usuario hacia las respuestas que queremos obtener, reforzando nuestro modelo mental.

Una vez que tenemos los datos, los algoritmos los usarán para modelar una situación del mundo real. Aquí nos encontramos con la segunda posibilidad de sesgo: por muchos datos que puedan manejar, es imposible tenerlos todos en cuenta y hay que hacer una simplificación previa, ya sea eliminando datos que se consideren poco relevantes o sustituyendo datos que sean difíciles de obtener por otros que consideramos equivalentes pero son más sencillos de obtener. Por ejemplo, podemos asociar la inteligencia de una persona a las notas que ha sacado en una asignatura concreta. Es algo muy fácil de obtener, pero apenas refleja la realidad de esa persona. Si ha sacado malas notas puede ser por otros motivos diferentes: problemas familiares, factores externos, que haya preferido abandonar esa asignatura temporalmente para centrarse en otras... esa información es mucho más complicada de obtener, por lo que se tiende a simplificar, a buscar un dato equivalente que valga como referencia aunque sea menos exacto.

Igual que simplificamos los datos de entrada, es muy tentador simplificar los datos de salida a un único valor, a un número, a una puntuación, como el resultado de un examen: "El oráculo ha hablado y ha indicado que eres apto. Que no te concedemos el crédito. Que esta persona es tu media naranja". Nuestra suerte queda sellada y poco podemos hacer contra ella. El proceso es una caja negra que no sabemos cómo funciona y que entrega un veredicto que no podemos recurrir. Si deciden que nuestro currículo no es válido o que no nos conceden la hipoteca, no podemos saber qué ha pasado: si hay un dato incorrecto, si se ha mal interpretado algo, si nos hemos quedado lo suficientemente cerca para volver a intentarlo en la próxima convocatoria, ...

Los bancos, los buscadores, las aseguradoras, nos reducen a un número que puede tener repercusiones importantes en nuestra vida. Sin embargo nosotros no podemos someter a ese mismo escrutinio a los algoritmos

Ese algoritmo, esa caja negra a la que hemos dado acceso y se nutre de nuestros datos más íntimos y privados, es un bastión inexpugnable protegido por las leyes de propiedad intelectual. Sin embargo, usa nuestros datos para diseccionarnos, para aprender nuestras motivaciones, para someternos a escrutinio, sin revelar nada a cambio. Esto provoca una situación de desigualdad: la privacidad personal va desapareciendo a la vez que la privacidad empresarial se fortalece. Los bancos, los buscadores, las aseguradoras, nos reducen a un número que puede tener repercusiones importantes en nuestra vida. Sin embargo nosotros no podemos someter a ese mismo escrutinio a los algoritmos: muchas veces ni siquiera somos conscientes de que hay un algoritmo detrás, ni de las consecuencias que puede tener sobre nuestra vida. No olvidemos que detrás de los algoritmos están las empresas que los han creado y que el objetivo de una empresa es ganar dinero. Así, podemos ver cómo el orden de los resultados de los buscadores o de las webs que recomiendan productos o servicios dependen del dinero invertido por los anunciantes: algo que resulta obvio pero que muchas veces olvidamos.

Por todo esto, es necesario que los programadores, los analistas de datos, los estadísticos que participamos en la elaboración de estos modelos hagamos un examen de conciencia y comprendamos nuestros propios

sesgos. Que analicemos los datos, cómo los obtenemos y cómo los simplificamos, para asegurarnos que son unos datos diversos e inclusivos, no centrados en un subconjunto de la población al que ¡oh, casualidad! pertenecemos y que nos hace pensar que el resto de personas también pertenecen. Debemos ir un paso más allá y analizar qué resultados se obtienen y qué se va a hacer con esos resultados, para ponernos en la piel de la persona que está al otro lado del algoritmo, la persona que recibe la valoración o la recomendación y cuyo futuro puede depender de unas decisiones que tomamos sin analizarlas lo suficiente.

Eubanks, V. (2018): *Automating Inequality* (edición de Kindle). Nueva York, St. Martin's Press.

Ferguson, A. G. (2017): *The Rise of Big Data Policing* (edición de Kindle). Nueva York, NYU Press.

O'neil, C. (2016): *Weapons of math destruction* (edición de Kindle). Nueva York, Crown Publishers.

Pasquale, F. (2015): *The black box society* (edición de Kindle). Cambridge, Harvard University Press.

Umoja Noble, S. (2018): *Algorithms of Oppression* (edición de Kindle). Nueva York, NYU Press.

Wachter-Boettcher, S. (2017): *Technically Wrong* (edición de Kindle). Nueva York, W. W. Norton & Company.