

Auditar la sociedad algorítmica

Gemma Galdon Clavell, presidenta de Eticas Foundation

A medida que los algoritmos y las soluciones de Inteligencia Artificial (IA) se generalizan, es urgente hablar de cómo proteger a la sociedad de los impactos de sistemas de datos que toman decisiones sobre nuestras oportunidades, derechos y posibilidades vitales. Si un día te niegan la prestación de desempleo, la atención médica o un contrato porque lo dicta un algoritmo, ¿a quién acudirás?

Es ya casi extraño abrir un periódico y no encontrar alguna referencia a los algoritmos. Ya sea en la sección de economía, tecnología o estilos de vida, estos procesos matemáticos se han convertido en poco tiempo en una pieza fija de nuestra cotidianidad informativa.

Los algoritmos se utilizan, por ejemplo, cómo primer filtro de candidaturas que aspiran a un lugar de trabajo. En un número creciente de lugares, también para determinar a qué prestaciones tiene derecho la ciudadanía y en qué cuantía, o para asignar riesgo y priorizar la atención social —a mujeres maltratadas, a menores excluidos, a personas sin hogar— o médica —en el triaje en la sala de urgencias, en las listas de espera para intervenciones quirúrgica—.

En un mundo donde todas las personas y procesos generan datos de forma permanente, la capacidad de los algoritmos de procesar estos datos masivos en tiempo real es la razón por la que hablamos cada vez más de ellos. Un algoritmo bien “entrenado” puede identificar un movimiento sospechoso en una tarjeta de crédito sin ninguna intervención humana: solamente analizando todas las instancias anteriores en que movimientos similares han resultado en fraude, por ejemplo.

En los últimos años, diferentes procesos tecnológicos nos han permitido avanzar en lo que se conoce como “aprendizaje automático”, que permite que los algorit-

mos tomen decisiones no sólo en base a los datos con los que se les ha entrenado, sino que incorporen la información que van recogiendo y generando en tiempo real. Con el aprendizaje automático, por ejemplo, un algoritmo de traducción automática será capaz no sólo de aplicar los datos (textos traducidos) con los que fue programado, sino de aprender de todos los textos que vaya analizando y traduciendo para ir mejorando de forma autónoma. Esta autonomía es la que alimenta los ríos de tinta que se dedican a la potencial sustitución de las personas por la Inteligencia Artificial.

Sin embargo, en el mundo real la aplicación práctica de los algoritmos en contextos de impacto social ha mostrado ya su cara más pernicioso. Por ejemplo, cuando Microsoft lanzó un perfil de Twitter de IA para demostrar como éste podía interactuar con el resto de usuarios de formas cuasi humanas, sus promotores se dieron de bruces con la inescapable realidad de que capturar el entorno y reproducirlo de forma autónoma no es algo necesariamente deseable: en pocas horas ese perfil empezó a tener opiniones racistas y nazis, ya que esto es lo que le “alimentaron” el resto de usuarios de la red social.

Igual que ese perfil de Twitter fue incapaz de entender qué ocurría a su alrededor y gestionarlo de forma responsable, muchos de los algoritmos que nos rodean se presumen neutrales mientras capturan y repro-

“Capturar el entorno y reproducirlo de forma autónoma no es algo necesariamente deseable”

ducen dinámicas sociales perniciosas o ilegales, desde la discriminación de género a la exclusión basada en el lugar de residencia, la edad o cualquier pieza que no encaje con la normalidad algorítmica. Recientemente ha sido noticia la tarjeta de crédito de Apple Card, que puede dar una línea de crédito 20 veces superior a un hombre frente a una mujer con mejor historial financiero. Ante estas derivas, ¿cómo garantizamos que la inteligencia artificial no amplifique lo peor de las dinámicas de discriminación y exclusión?

La discriminación de género algorítmica como la que aplica Apple es consecuencia de que, en el proceso de entrenamiento de las herramientas de IA, a ninguno de los diseñadores (seguramente todos hombres) se les ocurrió corregir este sesgo de la muestra, ni incorporar mecanismos de mitigación. Tampoco nadie les pidió que lo hicieran. La desregulación del espacio algorítmico es tal que no existen aún espacios de escrutinio sobre cómo estos sistemas toman decisiones o qué derechos tienen aquellos que se someten a ellas. La cascada de noticias sobre las víctimas de los procesos algorítmicos está poniendo sobre la mesa la necesidad de entender y regular estas nuevas realidades tecnológicas a través de prácticas concretas que trasciendan a la invocación de principios generales y los compromisos de buenas intenciones.

Si los algoritmos quieren capturar procesos sociales e incidir sobre oportunidades vitales, quienes los desarrollan deberán demostrar que cuentan con herramientas para comprender la complejidad de la sociedad, atravesadas por prejuicios, sesgos e injusticias que no deben reproducirse y amplificarse en sistemas algorítmicos. Recientemente un equipo de investigadores descubrió que el algoritmo de priorización de la atención médica en Estados Unidos asignaba a los pacientes afroamericanos menos

prioridad que a los blancos frente a dolencias parecidas precisamente porque en el diseño del sistema no se tuvo en cuenta que los colectivos minoritarios tienden a desconfiar más de los sistemas de salud (por motivos históricos y culturales) y a recibir peor atención cuando acuden a ellos. Con estos datos históricos, los ingenieros diseñaron un sistema de *perpetuación y ampliación* de esa desatención, a pesar de que la existencia de estos sesgos en la atención médica está ampliamente estudiada.

A pesar de lo terriblemente preocupantes y generalizadas que son estas malas prácticas algorítmicas que desatienden la mitigación de sesgos, los impactos negativos de estos sistemas pueden ir mucho más allá. Cuando el sistema escolar de Washington DC incorporó un algoritmo para decidir a qué profesores despedir, decidió construir su algoritmo en base a los resultados de sus alumnos en dos asignaturas concretas (Matemáticas y Lengua). Este intento de objetivación algorítmica del despido, sin embargo, acabó rescindiendo contratos a docentes con resultados excelentes en otras asignaturas, y generando incentivos perversos en la plantilla restante, que dejó de tener motivos para esforzarse en la impartición de asignaturas que no fueran las que recogía el algoritmo.

Un algoritmo, además, puede estar sesgado desde su propia conceptualización. ¿Por qué se centra la sociedad algorítmica en el control de los pobres y los trabajadores? Varios autores han denunciado que los desarrollos tecnológicos actuales parecen estar creando una sociedad dual en la que los ricos cuentan con el privilegio de una atención personalizada, humana y regulada mientras que a los grupos vulnerables se les condena al *Far West* de la máquina, en el que no existe transparencia, derechos ni procedimientos claros para apelar las decisiones algorítmicas.

Muchas de estas cuestiones merecen una reflexión social profunda y la articulación de consensos sociales sobre qué es y qué no es aceptable y deseable. Pero para empezar a articular estos consensos hacen falta herramientas que nos ayuden a entender qué es lo que está ocurriendo. ¿Pueden las auditorías algorítmicas ser esa herramienta?

Auditar la sociedad algorítmica tiene sen-



“Auditar la sociedad algorítmica tiene sentido si supone un ejercicio de escrutinio e incorporación de mejoras por entidades externas”

tido siempre que estas auditorías constituyan un ejercicio real de escrutinio e incorporación de mejoras realizado y verificado por entidades externas independientes. Utilizando procesos estandarizados y obligatorios que evalúen no sólo desde una perspectiva técnica, que nos condenaría a dar por válido un algoritmo discriminatorio siempre que la realidad también lo fuera; sino desde el cuestionamiento mismo de las hipótesis sobre las que recogemos los datos y del estudio de sus impactos concretos. Auditorías que valoren no sólo los procesos de datos, sino también las dinámicas que permean esos datos, las relaciones de poder y las desigualdades que condicionan y dan forma a nuestras sociedades. Que planteen de forma transparente qué sesgos existen, a qué grupos hay que proteger y de qué formas, y que velen por la correcta implementación de las decisiones algorítmicas

hasta la última milla del impacto social: trabajando con usuarios y ciudadanos/as para entender cómo les afectan los sistemas que se diseñan y como proteger a las víctimas de los falsos positivos (casos en los que un algoritmo se equivoca), velando por la existencia de mecanismos de apelación y restitución de los derechos mermados, y fomentando la rendición de cuentas de los sistemas de datos.

La auditoría algorítmica puede permitir mejorar el encaje legal de estos sistemas, a los que la ley exige “explicabilidad”, pero también mejorar las tecnologías que se desarrollan. Si nos parece inconcebible un coche sin límite de velocidad ni cinturones de seguridad, un medicamento sin descripción de efectos secundarios o un alimento sin etiqueta, ¿por qué aceptar algoritmos sin garantías?

¿Por qué no competir en la carrera tecnológica con sistemas desarrollados por los mejores profesionales técnicos, pero también con equipos excelentes en la comprensión de las dinámicas contextuales relevantes y de las especificidades de la prestación de un servicio en concreto? Si estos procesos matemáticos capturan datos y toman decisiones de forma cada vez más autónoma, ¿por qué no exigir la existencia de procesos independientes de certificación de prácticas e impactos? ¿A quién dan miedo las auditorías algorítmicas? ●