La inteligencia artificial se está expandiendo rápidamente por su promesa de eficiencia y ahorro de tiempo en múltiples procesos. Pero todavía no es perfecta. Diariamente, decisiones automatizadas invisibles influyen en la vida de las personas. Existen algoritmos para determinar si una persona merece la libertad condicional, si una víctima de violencia está en peligro o si alguien debe obtener una pensión de bienestar social. Y todos ellos se equivocan a veces, con consecuencias graves, incluso letales.

[ILUSTRACIÓN: MOOR STUDIO/ ISTOCK]

La herramienta algorítmica *VioGén*[1] se utiliza en las comisarías españolas desde el 2007. Sirve para evaluar el riesgo que corre una víctima de violencia machista de ser atacada de nuevo por su agresor, y pretende ser una ayuda para la policía y la justicia pero no es precisa, y se equivoca. En 2024, The New York Times investigó *VioGén* y denunció que 247 mujeres evaluadas con este sistema habían sido igualmente asesinadas. A 55 de las fallecidas se les había adjudicado un riesgo "bajo" y no se les dio ninguna protección. El sistema *VioGén* ha evaluado hasta la fecha 92.000 casos, consiguiendo que la atención a las víctimas fuese más rápida y eficiente. Pero ello no evitó que, por un error en sus cálculos, muriesen esas 247 mujeres.

247 mujeres evaluadas con la herramienta *VioGén* habían sido igualmente asesinadas

Este no es un artículo para mostrar las excelencias de la IA –que las tiene y resolverá de manera eficiente un sinfín de cuestiones relacionadas con tareas repetitivas, lo que permitirá reducir costes y tiempo en empresas y administraciones públicas–. Aceptada esta premisa, es importante poner sobre la mesa el concepto de rendición de cuentas cuando la tecnología causa perjuicios a las personas de manera injustificada.

Decisiones que afectan a la vida de las personas

VioGén no es el único sistema que ha sido cuestionado en España en los últimos años. Otros dos ejemplos son *RisCanvi* –utilizado en las prisiones de Catalunya desde el 2009 para estimar la reincidencia criminal– y el algoritmo del bono social eléctrico, Bosco –utilizado por las empresas eléctricas para decidir quiénes lo reciben–. Un último caso cuestionado: la herramienta que la Seguridad Social utiliza para predecir la probabilidad de que una persona con baja laboral esté lista o no para ocupar su puesto de trabajo. Medios de comunicación, expertos en tecnología y derechos humanos han puesto en duda repetidamente los sistemas mencionados. Se reclama conocer públicamente dónde se aplican los algoritmos que dependen de un organismo público. Asimismo, es necesario que exista una reparación del daño en caso de fallo, que haya una actualización y mejora al demostrar un error, y que se «transparente» o publique el código.

¿Cómo sabemos que la decisión es de un algoritmo?

Lo terrible es que no lo sabemos. Después del reportaje del The New York Times, el Ministerio del Interior presentó el *VioGén 2*, una versión revisada que promete calibrar mejor el riesgo de reincidencia en la agresión. En este punto, surge una pregunta: ¿por qué la policía y los jueces no informan a las mujeres agredidas que utilizan este sistema para decidir la protección que le darán? Si no sabemos, tampoco podemos exigir responsabilidades.

Las decisiones automatizadas siempre han de ir acompañadas de la valoración de un profesional

Según una auditoría externa del sistema *VioGén* realizada por la Fundación Éticas, uno de sus errores "es que el sistema no considera factores contextuales, como el historial completo de violencia, el miedo no verbalizado y la dependencia económica o emocional".

Cuando las aplicaciones algorítmicas pueden afectar seriamente la vida de las personas, como en el ámbito de la salud, el judicial (prisiones, sanciones o condenas) o el social (conceder ayudas a colectivos necesitados), las decisiones automatizadas siempre han de ir acompañadas de la valoración de un profesional. Así lo explicita la Autoridad Catalana de Protección de Datos en su informe *Inteligencia artificial, Decisiones Automatizadas en Catalunya*. Un doctor puede contar con la ayuda de una herramienta en la detección del cáncer, pero no se debería prescindir nunca de su valoración profesional para validar el diagnóstico, detallan en el informe.

Tener en cuenta los sesgos

Por otro lado, aún suponiendo que supiéramos que se está aplicando un sistema algorítmico cuando un policía nos toma declaración, hay que prever el riesgo de sesgos. Si el algoritmo da una decisión diferente a la del profesional experto (sea el funcionario de prisiones, en el caso del RisCanvi, o el policía, en el caso de VioGén), ¿cuál pesa más en la balanza? ¿Qué decisión se aplica finalmente?

La Vanguardia publicó un artículo donde indicaba que, desde 2009, "solo en un 3,2 % de las 89.638 veces en que se había aplicado *RisCanvi*" el profesional de prisiones había llevado la contraria al algoritmo.

Nos falta transparencia

Las reclamaciones de la ciudadanía por una transparencia real de los sistemas algorítmicos son totalmente lícitas, en un momento en que Europa cuenta con el Reglamento de Inteligencia Artificial de la UE, aprobado en 2024. Queda todavía mucho por escribir y estandarizar cuando se habla de "transparencia algorítmica", pero podemos adoptar la definición de los investigadores Eyad Hakami y Davinia Hernández Leo: "la transparencia se considera un medio para comprender la verdad y los motivos detrás de las acciones".

Han empezado a aparecer en nuestro país los Registros de Algoritmos, al modo de los actuales Portales de Datos Abiertos obligados por la Ley de Transparencia

La nueva ley europea obliga a explicitar públicamente los sistemas que son considerados de "alto riesgo", sobre todo, cuando afecte a sectores críticos como la sanidad, la educación o la justicia. Vamos bien, por aquí. Pero esta medida es insuficiente porque solo se registrarán los considerados "de alto riesgo", dejando fuera muchos otros que también afectan derechos fundamentales", reclama la organización ciudadana Civio. "Pedimos prohibiciones sobre el uso de reconocimiento facial en espacios de acceso público y del reconocimiento de emociones en todas sus formas", exigen en Algorights, una red colaborativa enfocada en los usos de la IA por parte de la administración pública.

Es justo mencionar los registros de algoritmos que, al modo de los actuales portales de datos abiertos obligados por la Ley de Transparencia, han empezado a aparecer con cuentagotas y de manera modesta. Pero son pocos y no hay un estándar europeo para que finalmente se puedan comparar sistemas de IA entre países. En España, hay que mencionar el registro de la Comunidad Valenciana, al País Vasco, que está empezando y a Catalunya, que publica fichas técnicas detalladas con información técnica y aspectos éticos a considerar.

Organizaciones sin ánimo de lucro llevan años pensando en cómo ser realmente transparentes en el uso de la IA y proponen diseños como el elaborado por IA Ciudadana, una coalición de organizaciones independientes.

Mientras, los gobiernos están en una carrera desenfrenada por encontrar maneras de resolver con la tecnología los problemas actuales. Pero, como dijo el Relator de Inteligencia Artificial y Privacidad del Consejo de Europa, Alessandro Mantelero: “Si digo que invertiré mucho dinero para predecir y controlar el crimen con una aplicación de IA policial, la pregunta ética y social podría ser: ¿por qué no invertir en un buen sistema educativo que intente reducir la criminalidad y la diferencia social? Eso atacaría el problema de origen”.

Responsabilidad algorítmica

Europa será recordada por sistemas algorítmicos tan fatídicos como el aplicado en la ciudad holandesa de Rotterdam que acusó erróneamente de fraude a miles de familias migrantes; o como el que dejó fuera de la universidad a estudiantes de barrios pobres del Reino Unido injustamente. Los dos ejemplos se mencionan porque una vez denunciados, los respectivos gobiernos fueron capaces de reconocer el error y retirar el sistema de IA.

¿Cuál es el objetivo del algoritmo? ¿Qué administración es la responsable? ¿Qué supervisión humana hay detrás del resultado? ¿Cómo se han mitigado los posibles sesgos? ¿Qué decisiones automatizadas toma y qué impacto puede tener en las personas? ¿Con qué datos se ha entrenado el modelo? Estas son solo algunas de las cuestiones principales que la administración pública debería explicitar a la ciudadanía. En un lenguaje sencillo, pero con suficiente detalle técnico en caso que se quiera explorar más en profundidad.

Hasta ahora, la reparación del daño solo se ha conseguido después de investigaciones periodísticas en profundidad y con fuertes movilizaciones ciudadanas. ¿No sería más fácil si se ejerciera una responsabilidad pública cuando los sistemas algorítmicos afecten negativamente la vida de las personas? La ciudadanía tiene derecho a exigir una reparación cuando la tecnología falla. La automatización de las decisiones no puede ser una excusa para diluir la rendición de cuentas. Como dice el Director del Instituto de IA del Barcelona Supercomputing Center, Ricardo Baeza-Yates: “La IA responsable no es un término de moda: es un compromiso para hacer que el mundo sea mejor lugar”.