¿Pueden los algoritmos ser más justos que las personas?

Por: Karma Peiró, Ricardo Baeza-Yates. 15/02/2022

¿Qué grado de confianza concedemos a las máquinas y qué consideración tenemos hacia la experiencia de los profesionales?

Las inteligencias digitales se utilizan cada vez más para resolver cuestiones complejas que antes se decidían a partir de la experiencia de los profesionales. Analizamos, pues, de qué manera toman decisiones los humanos y las máquinas y quién lo hace de forma más justa según el contexto.

Aixa es una niña de cinco años que llega a Urgencias de un hospital acompañada de su abuela. La menor presenta claros síntomas de maltrato y contusiones por todo el cuerpo. La mujer explica a la doctora que fue golpeada por sus padres. Para deducir si existe riesgo de reincidencia y evitarle nuevas palizas a la niña, la doctora plantea preguntas sobre el escenario de los hechos: ha de decidir entre derivarla a los servicios sociales para que sea protegida judicialmente o devolverla a su hogar, donde podría repetirse el maltrato. Para ello, cuenta con la ayuda de una herramienta de inteligencia artificial (IA) que estimará la gravedad del caso y las medidas cautelares a partir de las respuestas obtenidas. Si la decisión final es que ha de ser separada de los padres, Aixa no regresará a casa esa noche, pasará por centros de menores y casas tuteladas, donde adultos desconocidos se harán cargo de sus necesidades básicas durante los próximos años y hasta que alcance la mayoría de edad.

Cuando en 2018 la escritora y matemática Cathy O’Neil visitó el CCCB para presentar su libro *Armas de destrucción matemática*, lanzó una pregunta al público: «¿Quién sería capaz de decidir si una menor que ha sido maltratada debe o no regresar a su hogar, con la buena intención de protegerla de futuros golpes?». La audiencia permaneció muda, sin saber qué responder. «En Estados Unidos, un algoritmo lo está decidiendo», añadió. En ese momento, seguro que entre el público muchos se preguntaban: Pero… ¿cómo puede saber un algoritmo si la menor va a recibir otra paliza antes de que ocurra? Y segundos más tarde, la propia O’Neil respondió: «No lo puede saber, pero está ocurriendo».

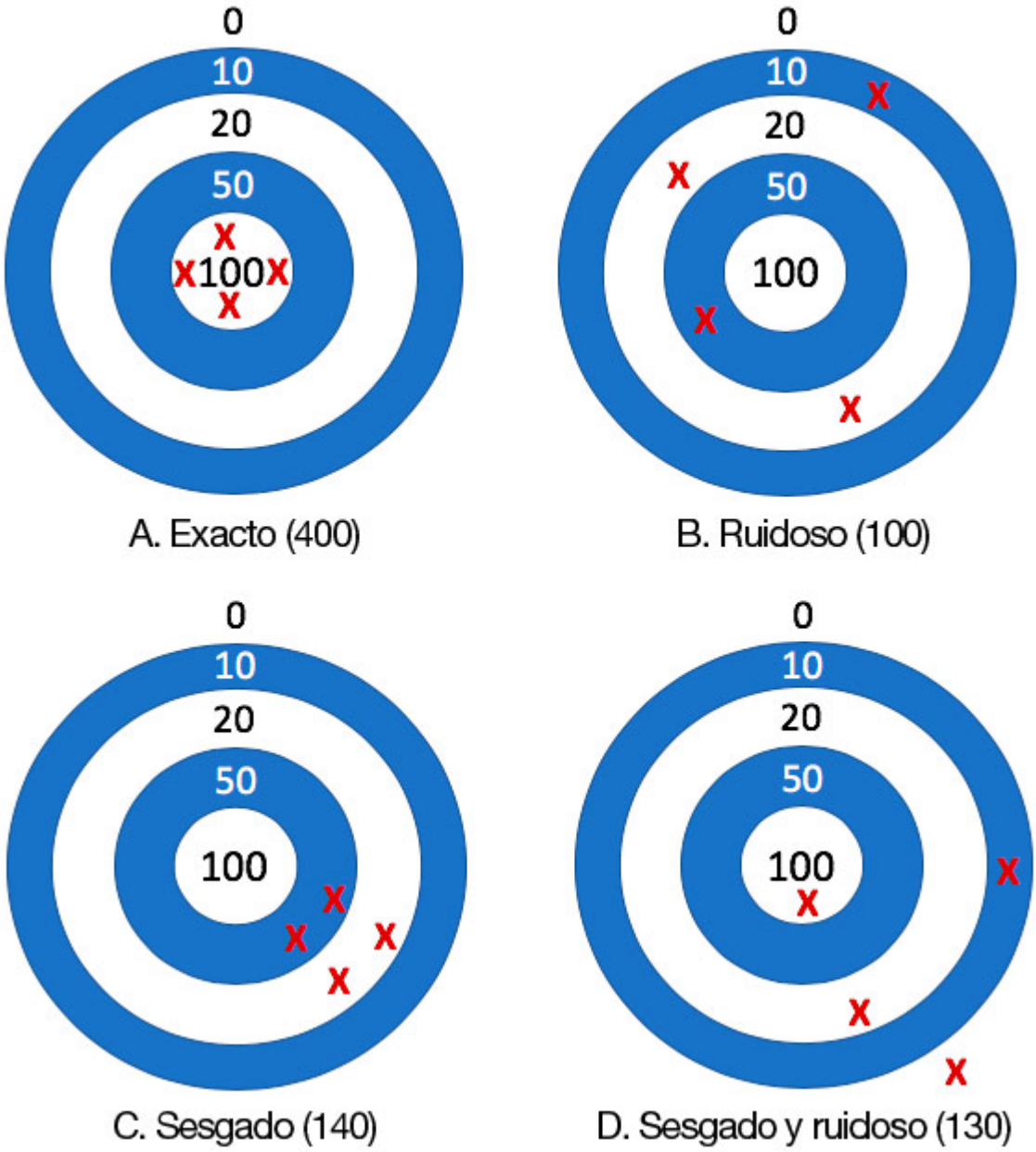
Hace tres años la pregunta podría haber parecido provocativa y lejana, pero actualmente este tipo de herramientas operan en la mayoría de los países sin que la población sea consciente de ello. Los profesionales que llevan casos sociales toman a diario decisiones complejas sobre asuntos muy diversos. Antes del *boom* de la IA, se basaban en la experiencia propia acumulada durante años.

Obviamente, se equivocaban en sus decisiones, lo que podía generar graves consecuencias para las personas afectadas. Por ello ahora se espera que los sistemas inteligentes de IA nos ayuden a mejorarlas. Pero ¿qué sucede si la herramienta de IA ofrece un resultado diferente al del especialista? ¿Cuál prevalecerá? Si el algoritmo decide que Aixa corre un riesgo alto y la doctora no lo considera tan grave como para separarla de su núcleo familiar, ¿se atreverá a enviar a la niña a casa?

Sabemos que la IA no es perfecta y que hasta el mejor sistema inteligente comete errores. Por tanto, cabe preguntarnos: ¿Qué grado de confianza se le concede a la máquina y qué confianza se tiene en la experiencia del profesional? Para dar una respuesta, tal vez debiera analizarse cuál de las dos decisiones –la humana o la de la máquina– tiene más sesgo y más ruido (es decir, mayor varianza en las decisiones). Esto dejando de lado la tendencia humana a aceptar las recomendaciones de un sistema que creemos inteligente.

Diferencias entre sesgo y ruido

El economista Daniel Kahneman escribió en 2016 –junto a otros autores– un artículo titulado «Ruido: Cómo superar el alto y oculto coste de la toma de decisiones incoherentes». En él se explicaba la diferencia entre «sesgo» y «ruido» haciendo una analogía con el resultado de lanzar dardos a cuatro dianas. La diana A no tiene ni sesgo ni ruido; la diana B tiene solo ruido; la diana C, solo sesgo; y la diana D tiene ambos. El diagrama siguiente se basa en esta idea.



Pero empecemos por el principio: ¿Qué se entiende por «sesgo» y qué por «ruido»? La IA es buena para establecer una gran cantidad de patrones y relaciones, así como para agilizar procesos y operaciones con datos masivos (*big data)*. Sin embargo, los algoritmos no son neutros. Tampoco los datos que sirven para entrenarlos, porque tienen sesgos. No hay duda de que los algoritmos de IA presentan sesgos, pero los seguimos utilizando porque el beneficio o el acierto de los resultados es considerablemente superior (en más de un 90%) al perjuicio o al

error. Los sesgos son parecidos a los prejuicios: todos los tenemos, en menor o mayor grado. Muchos los heredamos de nuestro entorno social o familiar sin darnos cuenta. El sesgo más grande es creer que no tenemos ningún prejuicio. Pero… atención: si los sesgos no se corrigen, corremos el riesgo de habitar un futuro en el que el progreso social sea cada vez más difícil a causa de la perpetuación de los prejuicios.

Por otro lado, el concepto «ruido» podría definirse como las diferentes decisiones que se pueden tomar respecto a una acción (individual o colectiva) en una situación personal o profesional. Cuando se pidió a los desarrolladores de *software* de una empresa que estimaran el tiempo que tardarían en desempeñar una tarea determinada, las horas contabilizadas variaron de media en más de un 70%. ¿Por qué, si la tarea es la misma para los mismos perfiles de profesionales de una misma empresa? A menudo se toman decisiones que se alejan de las que tomarían otros colegas, de las propias tomadas anteriormente o de las que uno mismo afirma seguir.

Llegados a este punto, es bueno saber que la IA tiene sesgos, pero no ruido. En cambio, a las personas nos influyen ambas variables.

El ruido es fácil de percibir

Regresemos ahora a las dianas. Si las viéramos por detrás, nos daríamos cuenta de que B y D tienen una mayor variabilidad, y notaríamos el ruido fácilmente. Pero entre A y C no es tan fácil saber cuál es la diana con sesgo. Es decir, el ruido es fácil de percibir, mientras que no siempre somos conscientes de los sesgos. La mayoría de las decisiones humanas tienen ruido, por eso buscamos una segunda opinión en cuestiones médicas o deseamos que el juez o jueza que nos asignen esté de buen humor el día de nuestra audiencia. Estos son dos ejemplos de ruido a nivel individual, pero nosotros tampoco somos consistentes a lo largo del tiempo.

En 2007, investigadores de Stanford analizaron las decisiones de peticiones de asilo en EE. UU. y hallaron que un juez aprobaba el 5% de ellas, mientras que otro aprobaba el 88% de las peticiones. Los dos tienen un sesgo muy distinto, pero el sistema judicial en su totalidad aparece como ruidoso, porque las decisiones de sus jueces no son consistentes. Finalmente, el sesgo puede ser tanto negativo (machismo) como neutral (paridad de género). Al final, es una cuestión de consenso

social.

Y ¿cómo es más justa la IA?

Una IA sin ruido es más justa en el sentido de que ante dos casos iguales, su previsión es la misma. Es decir, si dos acusados cometen el mismo crimen, la condena será la misma. Pero muy pocas veces los casos son idénticos. El paradigma de la injusticia social con un algoritmo lo sacó a la luz ProPublica –medio estadounidense independiente– en 2016, tras realizar una investigación para demostrar el sesgo que aplicaba el algoritmo COMPAS en perjuicio de los acusados negros.

La IA no tiene ruido porque los algoritmos son determinísticos. Es decir, ante una entrada similar, el resultado será similar. Por ejemplo, un controlador aéreo puede dirigir el tránsito de aviones mejor o peor en función de la situación de estrés, del cansancio o de las circunstancias que alteren su estado de ánimo. En cambio, un sistema inteligente siempre proporcionará los mismos resultados sobre la misma situación. Esto no puede decirse de los jueces.

En cambio, la IA sí puede ser desigual porque aprende de los sesgos intrínsecos de los datos. En el análisis «Human Decisions and Machine predictions», el científico Jon Kleinberg –junto con otros colaboradores– explicaba que el sistema entrenado con datos de fianzas de Nueva York era racista y amplificaba ligeramente las decisiones del juez más racista; esto resultaba un poco sorprendente, pues la única información demográfica que empleaba era la edad del acusado. En otras palabras, el racismo estaba codificado en los antecedentes criminales y lo deducía a partir de ellos. Pese a esto, el algoritmo era más justo, pues no tenía ruido. Manteniendo la misma tasa de fianzas podía disminuir la criminalidad en un 25%, o, manteniendo el mismo nivel de criminalidad, podía otorgar fianza a un 42% de los inculpados que antes no la obtenían. Peor aún, tomando en consideración los crímenes cometidos, los jueces concedían la fianza a más de la mitad del 1% de los criminales más peligrosos.

Sin embargo, como diría la eticista Lorena Jaume-Palasí, la justicia es en realidad mucho más complicada. Podría darse el caso de que un juez o jueza haya dejado a alguien en libertad en cumplimiento de la ley, incluso cuando sospecha que las probabilidades de reincidencia criminal son elevadas. «La justicia va más allá de

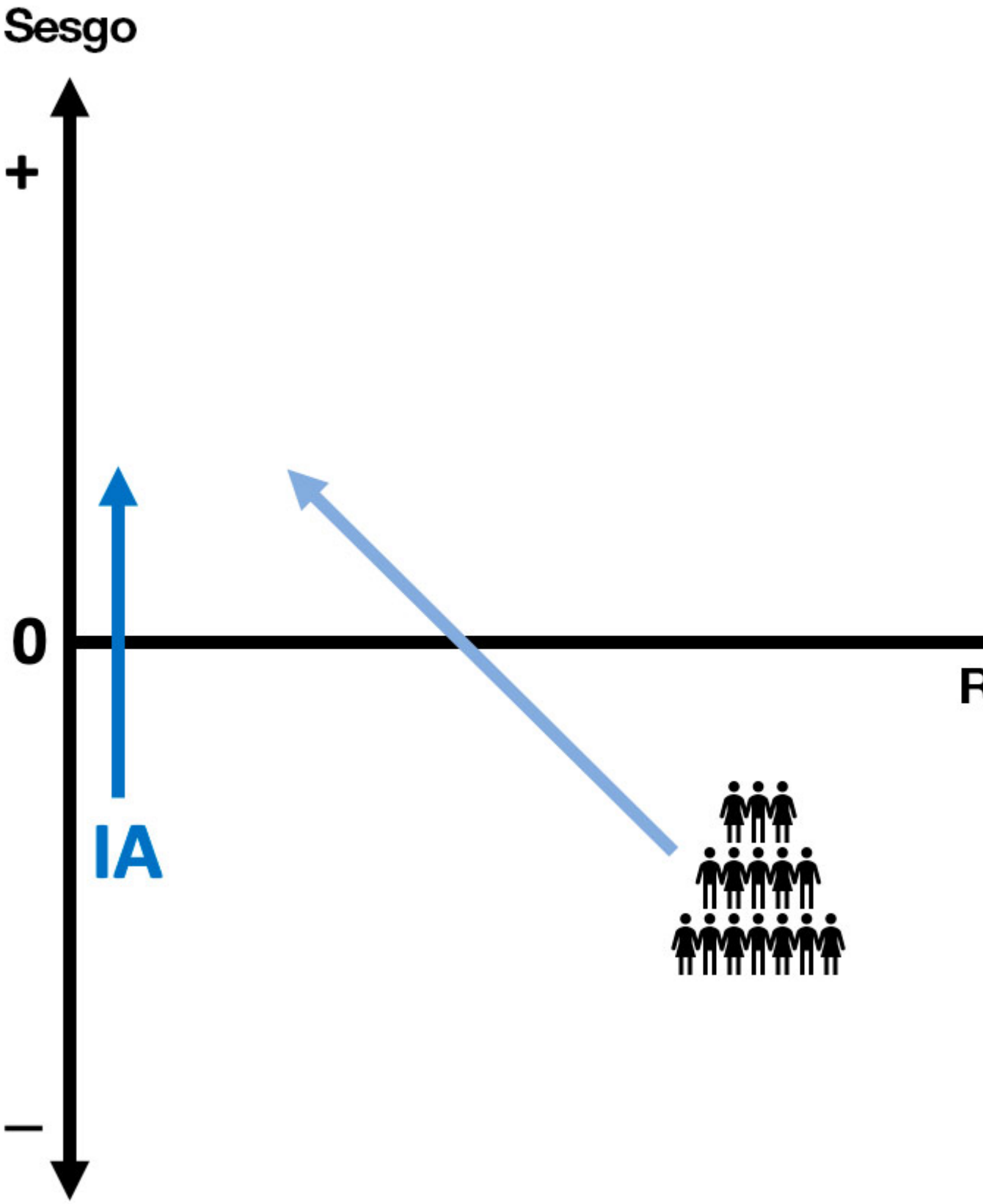
repetir el pasado a la hora de decidir un caso. Se debe contextualizar», explica Jaume-Palasí. «Pues aunque los antecedentes sean similares a otros casos, el contexto puede ser completamente diferente. ¿Es más justo alguien que discrimina de manera consistente pero no tiene en cuenta cada caso de manera individual? Si un sistema reinterpreta lo que requiere la ley, y los jueces emplean una metodología opuesta a la requerida, esto ya de por sí es injusto».

Sesgos positivos para la sociedad

¿Hemos de asumir que la IA siempre tendrá sesgos? No necesariamente. ¿Es posible que un algoritmo tome decisiones automatizadas sin sesgo para las personas o empresas? Eso dependerá de si los sesgos están en los datos, en el algoritmo o en la interacción entre el usuario y el sistema. Como en la práctica es realmente difícil que no haya ningún tipo de sesgo, lo que hay que hacer es mitigarlos. Pero será difícil eliminarlos, especialmente cuando no conocemos todas las fuentes. Los sesgos más importantes son explícitos, pero muchos otros pueden ser implícitos.

Los sesgos también pueden ser positivos, lo que significa que tienen un efecto favorable para la sociedad. Esta debería ser una condición exigible para cualquier algoritmo que tomara decisiones automatizadas o realizara predicciones. Por ejemplo, que haya un mayor número de enfermeras que enfermeros en los hospitales podría ser un sesgo positivo si quedara demostrado que ellas tienen más cualidades empáticas y de atención con los pacientes.

Con el tiempo, sería bueno que la IA ayudara a tener sesgos positivos (v. flecha azul vertical del diagrama adjunto), y que fuera tan útil como para que las personas percibieran sus propios sesgos a fin de reducirlos (v. flecha diagonal del diagrama). Pero… ¿no es pedirle mucho a esta tecnología? Si son las personas quienes programan la IA, y nosotros mismos no somos conscientes de nuestros sesgos, ¿cómo será posible que lo hagan por nosotros las máquinas? ¿No deberíamos, en primer lugar, cambiar todo nuestro sistema educativo global y aprender desde pequeños a identificar los sesgos sociales y cognitivos más relevantes?


Sesgo
+
0
IA
R
–

Más humanos con más tecnología

La IA despierta muchos recelos. Sus ventajas son sobradamente exhibidas pero también los riesgos que comporta por la multitud de errores identificados en los últimos años. Entre sus máximos retos está el de la explicabilidad, es decir, entender cómo un algoritmo ha aprendido de manera automática y presenta los resultados, ya sea en forma de decisiones automatizadas o de predicciones.

Otro riesgo –tal como menciona la artista e investigadora Joana Moll en su artículo «Contra la complejidad»– es que se nos haga creer que cualquier problema puede ser resuelto de manera sencilla por las máquinas. En este sentido, es inevitable preguntarse por las consecuencias que podría tener este *laissez faire* para cualquier sociedad. Si, por ejemplo, la doctora de Aixa confiara ciegamente en los resultados de los algoritmos que tiene a su disposición –desterrando su propia intuición derivada de la experiencia profesional–, los humanos dependeríamos totalmente de las máquinas para cualquier movimiento, avance o evolución.

Pero, sin ánimo de ponernos apocalípticos, ni de acabar con el peor de los escenarios, sería de esperar que la tecnología más avanzada nos ayudara a construir un futuro cada vez más respetuoso con los derechos y valores por los que hemos luchado en las últimas décadas. El cómo es todavía una incógnita. Ni siquiera sabemos si lo lograremos. Lo que sí cabría desear es que la IA fuese cada vez mejor y que complementase el conocimiento humano. Por ahora, no solo no corrige nuestros errores, sino que en ocasiones hasta los magnifica.Este articulo tiene reservados todos los derechos de autoría

LEER EL ARTÍCULO ORIGINAL PULSANDO AQUÍ

Fotografía: Lab.cccb

Fecha de creación
2022/02/15