Condenado por un algoritmo

José A. Estévez Araujo

El pasado mes de agosto fue liberado Michael Williams, un sexagenario que había permanecido durante un año en prisión preventiva en una cárcel estadounidense. Se le había acusado de asesinar a un joven al que recogió en su coche para acompañarle a casa durante las protestas de mayo del año pasado. La única "prueba" que había contra él era producto de un cálculo algorítmico. Un sistema de sensores especialmente diseñados para detectar y localizar disparos "registró" una detonación aquella noche y el algoritmo del sistema situó el disparo donde se encontraba Mr. Williams en ese momento. El acusado llevó al chico a un hospital y explicó que le habían disparado desde otro coche. La policía no encontró el arma del crimen ni pudo descubrir ningún motivo plausible para explicar que Mr. Williams hubiese matado al muchacho. No había testigos oculares ni prueba forense alguna.

Al cabo de unos meses, en agosto de 2020, unos policías fueron a buscar a su casa al Sr. Williams y, después de interrogarle, le pusieron a disposición del juez. La fiscalía solicitó que ingresara preventivamente en prisión y el magistrado le envió a la cárcel.

El sistema inteligente de detección de disparos de armas de fuego que se había instalado en la ciudad era gestionado por la empresa ShotSpotter. La compañía opera en 119 localidades de Estados Unidos, el Caribe y Sudáfrica. Es de destacar que la información proporcionada por los sensores no se envía directamente a la policía, sino que se remite al centro de datos de la compañía donde es analizada. Sólo después se da la alerta a la comisaría correspondiente. ShotSpotter no es la única empresa que ofrece este tipo de servicios.

Los sistemas automáticos de detección de disparos son objeto de fuertes críticas debido a los falsos positivos que generan. Parece, por ejemplo, que les resulta muy difícil distinguir las detonaciones de las armas de fuego de los fuegos artificiales. Pero el problema más grave que plantean estos sistemas es que la compañía mantiene en secreto los algoritmos que utiliza aduciendo que esa medida es necesaria para evitar que las empresas competidoras los copien.

El ocultamiento del programa fuente conduce a una restricción de los derechos de los acusados cuando los cálculos de ShotSpotter son utilizados como prueba inculpatoria en un proceso penal. Los actores en el juicio

desconocen qué tipo de datos se han registrado y cómo han sido analizados: el acusado, el juez, el jurado y los letrados no saben qué pasos se han seguido para generar la información que se presenta como prueba. Es algo parecido a lo que ocurrió durante la "caza de brujas" impulsada por McCarty: quienes comparecían ante la comisión de depuración desconocían las pruebas que existían contra ellos.

Los algoritmos se utilizan profusamente en el sistema sancionatorio estadounidense: sirven, entre otras cosas, para determinar si se concede o no la libertad condicional, si se suspende la ejecución de la pena o, incluso, se utiliza como criterio para fijar su duración. Todos estos algoritmos "calculan" la peligrosidad del acusado: el riesgo de reincidencia o de violar la condicional. Hay muchos abogados, activistas y organizaciones que luchan para que se desvelen los algoritmos que se utilizan en el proceso punitivo y se han conseguido bastantes victorias. Pero el desvelamiento del código fuente y de los modelos matemáticos que utilizan los algoritmos no constituye una garantía suficiente.

Algoritmos deductivos e inductivos

La opacidad de los algoritmos fue vivamente denunciada por Cathy O'Neil en un libro titulado *Armas de destrucción matemática*, reseñado por Ramón Campderrich en esta misma revista. Los peligros de la gobernanza algorítmica desvelados por la autora son muy serios y ya se están haciendo esfuerzos para intentar combatirlos.

Pero hay otros riesgos que no son contemplados por Cathy O'Neil, porque en su libro no distingue adecuadamente dos tipos de algoritmos muy diferentes, que podríamos denominar "deductivos" e "inductivos". Los primeros son series lógicas de instrucciones construidas sobre un modelo matemático de base que selecciona las características y las variables que van a ser tenidas en cuenta. Están compuestos de reglas en forma de sentencias condicionales: si se da la situación A, entonces haz B.

El funcionamiento de los segundos es analizado por Louise Amoore en un libro titulado *Cloud Ethics*, título que tiene el doble sentido de "ética en la nube" y "ética nublada". Esta autora se centra en la trascendencia ético-política de las decisiones que adoptan los algoritmos utilizados en un tipo de máquinas virtuales dotadas de inteligencia artificial: las redes neuronales. Estos mecanismos se diferencian de los algoritmos "deductivos" por dos características: a) son capaces de aprender y de reprogramarse autónomamente; b) los resultados a los que llegan son imprevisibles incluso para sus propios programadores.

Amoore recalca repetidamente la diferencia entre ambos tipos de algoritmos:

“La representación de los algoritmos como una cadena lógica pasa por alto el grado en que los algoritmos se modifican a sí mismos en y a través de sus relaciones iterativas no lineales con los datos de entrada” (p. 11).

"Entiendo que la escritura del algoritmo excede sustancialmente la escritura del código fuente y se extiende a la escritura iterativa, la edición y la reescritura de compuestos de datos, humanos y otros algoritmos" (p. 103).

Redes neuronales

Las llamadas “redes neuronales” son máquinas virtuales compuestas de “neuronas artificiales” conectadas entre sí. Las neuronas que integran la red tienen unas características funcionales similares a las que se encuentran en el cerebro humano. Pueden captar y emitir señales eléctricas. También tienen la capacidad de calibrar la importancia relativa de una señal atribuyéndole un determinado “peso”. Tienen asimismo un “umbral de activación” que determina cuándo emitirán una señal como reacción a un impulso concreto.

Las neuronas están distribuidas en diversas capas, dos de las cuales son “externas” y el resto se caracterizan como “ocultas”. La primera capa externa es la de “entrada”. Las neuronas que la integran reciben los estímulos de “fuera”: de los bancos de datos, de los sensores... La otra capa externa es la de “salida”. El output que proporciona esa capa consiste en la solución óptima al problema planteado y su probabilidad de éxito. Cada neurona de una capa está conectada con todas las de la siguiente. Las diferentes capas actúan como una especie de filtros. La red parte de una gran cantidad inicial de información, de variables y de posibilidades. Cada una de las capas ocultas va descartando unas soluciones y optando por otras. La de salida está diseñada para ofrecer una solución única. Como dice Amoore, la red neuronal lleva a cabo una “condensación” desde la multiplicidad hasta la unidad.

En el proceso de condensación, los algoritmos realizan innumerables decisiones entre posibles alternativas, en base a parámetros que ellos mismos establecen a partir de su capacidad de aprendizaje autónomo. Pueden modificar por sí mismos el “peso” que las neuronas dan a una determinada señal. [4]

Una red neuronal puede tener numerosas capas ocultas. En cada una puede haber una gran cantidad de neuronas. Hay redes que contienen millones de ellas. A mayor cantidad de capas y neuronas más complejos serán los problemas que podrá analizar la red. [4]

Relevancia ético-política de las decisiones algorítmicas

Amoore reflexiona en cada capítulo del libro sobre el funcionamiento de tipos diferentes de máquinas inteligentes: coches autónomos, robots médicos, sistemas de vigilancia policial que predicen disturbios, armas letales autónomas... Cada una de ellas le sirve para ilustrar distintos problemas ético-políticos que plantea su utilización.

Los algoritmos que Amoore analiza en cada uno de los capítulos del libro proporcionan respuestas "inmediatamente accionables". Dicen cosas como "hay una probabilidad muy alta de que este condenado reincida". Eso se puede traducir inmediatamente en imponerle una condena de cárcel en lugar de la obligación de realizar trabajos comunitarios, en alargar la duración de la pena de prisión o en negarle la libertad condicional. Lo mismo ocurre con los algoritmos que dicen "es altamente probable que esta persona incumpla los términos de su visado". La traducción en una acción consiste en negarle la entrada al país. Las decisiones de los algoritmos afectan a personas muy concretas. Pueden restringir sus derechos o incluso determinar si deben morir.

Las dos principales conclusiones del libro de Amoore acerca del funcionamiento y uso de los algoritmos que hemos llamado "inductivos" podrían formularse del siguiente modo: a) los algoritmos toman decisiones arbitrarias b) los algoritmos tratan a las personas como cosas.

Los algoritmos toman decisiones arbitrarias

Los algoritmos llevan a cabo ponderaciones al igual que hacen los jueces. En caso de conflicto, los órganos judiciales determinan qué derecho, principio o bien jurídico debe prevalecer. Las premisas y criterios utilizados para la ponderación pueden tener carácter no sólo jurídico, sino también ético o político. Los argumentos esgrimidos en las sentencias de los tribunales constitucionales para establecer que un derecho fundamental prevalece sobre otro lo ponen claramente de manifiesto.

La propia Amoore utiliza el término "ponderación" (*weighting*[3]) al referirse al modo como los algoritmos inductivos razonan:

"La disposición de las proposiciones hace que un resultado aparentemente óptimo surja de la *Ponderación* diferencial de los caminos alternativos a través de las capas de un algoritmo" (p.13)[4].

"Empezar por aquí es partir de la idea de que todos los algoritmos de aprendizaje automático siempre incorporan suposiciones, errores, sesgos y *ponderaciones* que son totalmente ético-políticas" (p. 75).

La diferencia entre los algoritmos y los jueces es que éstos últimos tienen que fundamentar sus sentencias. El juez tiene que determinar qué hechos se consideran probados y mediante qué pruebas. Ha de exponer los fundamentos normativos que le han llevado a dictar su fallo en relación con los hechos juzgados. El algoritmo nos da una solución y una probabilidad de éxito que sería equivalente al "fallo". Pero el usuario del algoritmo o el destinatario de sus decisiones desconocen cómo ha llegado el algoritmo a esa conclusión. Abrir la "caja negra" y "visualizar" el modelo matemático o el programa fuente no proporciona un conocimiento suficiente de los factores que se han tenido en cuenta ni de las valoraciones que se han llevado a cabo en el caso de los algoritmos que hemos denominado "inductivos". No nos dirá qué pesos y umbrales de activación han utilizado las neuronas. Desconoceremos de dónde han extraído la información y por qué han seleccionado unos rasgos de los datos y no otros.

El algoritmo nos proporciona la solución a un problema y su probabilidad de éxito (p. ej. un 90%). En su proceso de razonamiento, el algoritmo se encuentra con innumerables bifurcaciones. En diversos momentos puede haber escogido uno u otro camino basándose en una probabilidad menor (p. ej. del 60%). La probabilidad que da a su propuesta final oculta el grado de incertidumbre al que se ha enfrentado a la hora de realizar las opciones previas que finalmente le han conducido a proponer esa solución. Las decisiones de los algoritmos no están, por consiguiente, fundamentadas. No se exponen las premisas, valoraciones y opciones que han conducido a su output. Las decisiones de los jueces son recurribles por los afectados. Las de los algoritmos son inapelables.

Los algoritmos tratan a las personas como cosas

Los algoritmos tratan a las personas como meros conjuntos de atributos. Las consideran como elementos pertenecientes a diversas clases que, en ocasiones, el propio algoritmo ha creado. Estos conjuntos están definidos en forma intensiva, es decir, en base a las características que permiten determinar si un elemento forma o no parte del mismo. Facebook tiene clasificados a cada uno de sus usuarios en cientos de categorías diferentes. Hay plataformas que crean "gemelos digitales" generando archivos que contienen todos los datos sobre una persona. Pero en otros casos, no es necesario siquiera que todos los atributos de una persona sean asignados a un ente único. La persona como singularidad única e irrepetible no existe para el algoritmo. Cuando éste comete un error que perjudica a alguien, eso constituye para él únicamente una ocasión para aprender.

Los algoritmos se utilizan profusamente en el sistema sancionatorio

estadounidense como ya hemos visto. Algunos de ellos “calculan” la peligrosidad del acusado: el riesgo de que reincida o de que viole la condicional.

La determinación estadística de la peligrosidad es un método que castiga a determinadas personas, no por sus actos, sino por lo que pueden llegar a hacer en el futuro. Es lo que se ha caracterizado como una deriva “actuarialista” del derecho penal.

Pero, además de esto, hay otro aspecto importante. La peligrosidad de las personas se determina no en función de su trayectoria personal, sino en base a lo que han hecho otras personas en el pasado. Se les castiga no por su conducta, sino por la de otros.

Los bancos de datos proporcionan a los algoritmos información acerca de un enorme número de casos acaecidos en el pasado. En base a ellos se establecen correlaciones estadísticas entre determinados atributos (o combinaciones de atributos) y el peligro de reincidencia o de violación de la condicional. Por ejemplo, el algoritmo puede descubrir una correlación estadística significativa entre habitar en una determinada zona o formar parte de una familia monoparental y ser reincidente. Pero el sujeto acerca del cual se decide no debería ser hecho responsable de hechos pasados en los que la persona evaluada no ha tenido intervención alguna.

La necesidad de una reflexión “inteligente”

Existen multitud de trabajos que reflexionan acerca de la moral de los algoritmos y también propuestas institucionales acerca de cómo deben ser utilizados. Pero, desde mi punto de vista, muchas veces resultan insatisfactorias para resolver los problemas generados por los algoritmos que aprenden y se reprograman.

Los sistemas de inteligencia artificial plantean cuestiones de vida o muerte. Es el caso de los sistemas autónomos de armas letales. Pero las situaciones trágicas no se dan únicamente en el combate. Los coches autónomos también deben (o deberán) tomar decisiones que afectan a la vida y la integridad física de las personas ante la inminencia de un accidente.

Los algoritmos regulan cada vez más aspectos de nuestras vidas. Son utilizados por las empresas privadas y por los entes públicos. Interfieren con muchos de nuestros derechos fundamentales aparte del derecho a la vida. El derecho a la libertad en todas sus manifestaciones (de movimiento, de expresión, de reunión y manifestación...), el derecho a la intimidad y la

protección de datos, el derecho a la igualdad (concesión de ayudas estatales, autorización de un crédito...). Los algoritmos inteligentes son capaces de manipular nuestras emociones y condicionar nuestra conducta. La "expropiación de nuestro futuro" ha sido muy bien analizada por Shoshana Zuboff en su libro *La era del capitalismo de la vigilancia*.

La expansión de la gobernanza algorítmica puede ser comparada con la extensión de la burocracia a todos los ámbitos institucionales tanto públicos como privados. Weber resaltó la eficiencia, inigualable en su tiempo, de la organización burocrática. La burocracia podía utilizarse tanto para organizar un ejército como un hospital o una empresa de automóviles. Parecía un simple medio técnico para alcanzar un fin. La utilización ética de la organización burocrática dependería del objetivo al que sirviera. No obstante, poco a poco, se fueron manifestando las consecuencias estructurales de la burocratización del mundo. La organización burocrática carecía de mecanismos adecuados de retroalimentación. Su rigidez no le permitía adaptarse a las circunstancias cambiantes. La forma burocrática de tratar a las personas fue representada extraordinariamente bien por Kafka en sus novelas.

Ahora estamos viviendo un proceso de "algoritmización del mundo". Los algoritmos parecen ser la solución para todo. Es tal el entusiasmo que ya hay algoritmos que se han presentado a las elecciones como candidatos, prometiendo una gestión rigurosa e imparcial. Es necesario detectar las consecuencias estructurales negativas de ese proceso para prevenir o corregir los excesos perversos que tuvo la burocratización del mundo (y que el propio Weber ya previó).

El libro de Amoore es un paso en esa dirección. Forma parte de un corpus creciente de estudios críticos relacionados con la gobernanza algorítmica. Pero es un texto muy enrevesado y de difícil comprensión.

La autora utiliza de manera dogmática conceptos y planteamientos de Foucault y otros filósofos postestructuralistas. Intenta encuadrar sus planteamientos en el marco conceptual de estos autores. Sin embargo, resulta innecesario en su caso llegar a un plano tan abstracto. Desde mi punto de vista, hay un salto entre el grado de abstracción al que Amoore llega en el tratamiento de los problemas y el de las ideas y planteamientos de Foucault o Derrida que utiliza la autora. El proceso de abstracción y el proceso de concretización no llegan a ensamblarse. Existe un espacio vacío entre ambos.

En mi opinión, estos marcos conceptuales no hacen sino oscurecer innecesariamente la exposición. Con ello no cuestiono la fecundidad del pensamiento de estos autores. En el caso de otro libro relacionado con estos

temas, escrito por Orit Halpern[4]y titulado *Beautiful Data*. En él, la utilización del método genealógico de Foucault se muestra muy fecunda. La autora desvela el proceso, iniciado tras la segunda guerra mundial, que ha transformado la manera de percibir y comprender el mundo “datificado” de nuestros días por parte de las diversas disciplinas que se ocupan del comportamiento humano, desde la neurociencia hasta la sociología.

Pero en el caso de Amoore, insisto, me parece innecesario y cuestionable intentar encuadrar los problemas ético-políticos que plantean las decisiones algorítmicas en planteamientos foucaultianos. Por ejemplo, la autora pretende “solucionar” el enrevesado problema de la responsabilidad de los algoritmos negando la existencia misma del autor. Se trata de una forma de “echar balones fuera”: como no podemos determinar con claridad quién es el autor de las decisiones algorítmicas, cuestionemos el propio concepto de “autor”. Amoore incurre en una hilarante incongruencia cuando escribe: “Como señala Michel Foucault, la crítica y la filosofía tomaron nota de la desaparición —o muerte— del autor ya hace tiempo" (p. 91). Utiliza la “autoridad” de Foucault para cuestionar la existencia de autores propiamente dichos.

Es preciso profundizar el análisis crítico de la algoritmización del mundo, pero también es necesario hacer accesibles sus resultados a la ciudadanía. Contamos ya con la experiencia histórica de la toma de conciencia de los problemas ecológicos, con todas sus luces y sombras. El proceso de toma de conciencia de los problemas de la algoritmización en particular y de la digitalización en general debería ser mucho más rápido. Es necesario conseguir a tiempo que el debate público sobre la inteligencia artificial sea también “inteligente”.

29/9/2021