

What if? Algoritmos y ética

¿Y si los algoritmos se pudieran regir por principios éticos?

Traducción del documento

[http://www.europarl.europa.eu/RegData/etudes/ATAG/2018/624267/EPRS_ATA\(2018\)624267_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/ATAG/2018/624267/EPRS_ATA(2018)624267_EN.pdf)



Los algoritmos son simplemente pasos de cómo resolver un problema. Normalmente se representan a través de códigos informáticos, entendiendo estos como una serie de instrucciones que se deben seguir para completar una tarea.

Un algoritmo puede ser codificado por un programador o generado automáticamente a partir de datos, como en el aprendizaje de máquinas. Esta última opción se considera inteligencia artificial. Las decisiones que se toman día a día en el mundo están basadas cada vez más en técnicas de computación impulsadas por máquinas que aprenden algoritmos. Por ejemplo, las plataformas intermediarias que proporcionan alojamiento (AirBnB) o servicios de transporte (Uber) funcionan gracias a los algoritmos. Al mismo tiempo, los algoritmos de manera implícita o explícita no son neutros, ya que su diseño puede estar condicionado por la raza o el sexo de su programador. Esto plantea una pregunta importante: ¿es posible garantizar que los algoritmos sean éticos?

Los algoritmos se utilizan para tomar decisiones de manera automática con un gran impacto en los individuos y en la sociedad. Tienen la capacidad, entre otras cosas, de influir en la distribución de los beneficios sociales, bienes como la educación, el empleo, la protección policial y atención médica, así como en la protección de derechos fundamentales como el derecho a la vida, a un juicio justo, a la educación, a la presunción de inocencia, el derecho a la intimidad, la libertad religiosa, los derechos de los trabajadores. Casos como el del algoritmo Volkswagen, que permitía a los vehículos pasar las pruebas de emisiones mediante la reducción de sus emisiones de óxido de nitrógeno durante las mismas, o los programas informáticos policiales utilizados para localizar a los delincuentes reincidentes, que con frecuencia parecen estar sesgados contra la gente negra, son ejemplos de cómo la ética de los algoritmos ha conducido a "tragedias algorítmicas".

* Alumno en prácticas de la Universidad Pontificia de Comillas

El aprendizaje de algoritmos puede reproducir y agudizar los prejuicios que entablan algunas formas de clasificación social, las cuales manifiestan una nueva forma de “discriminación racial”, con el consecuente impacto en la vida de las personas. De hecho, los algoritmos pueden ampliar estos prejuicios ya que, indirectamente, contienen también los valores de sus programadores, quienes deciden qué tipo de datos incluir o excluir y cómo calcular cada componente. Si no se aplican medidas de auditoría o comprobación, los datos sesgados o incompletos pueden conllevar a modelos estadísticos fallidos y reforzar los prejuicios sociales. En lugar de erradicar los prejuicios humanos, los algoritmos podrían provocar que se magnifiquen o que se enquisten, lo que, a su vez, puede inducir que el ser humano pierda autonomía y soberanía, especialmente si estos algoritmos son capaces de crear nuevos algoritmos.

Así, un análisis de ProPublica 2016 sobre el uso de *software* automatizado de apoyo a la toma de decisiones bajo el nombre de El COMPAS¹ advirtió de los prejuicios raciales en el sistema de justicia penal de los Estados Unidos, dado el sistema empleado, el *scoring*. A pesar de ello, los jueces estadounidenses están utilizando cada vez más este algoritmo de evaluación de riesgos para fundamentar sus decisiones en algunas etapas del proceso de justicia criminal. En el caso de los datos de salud, existen tres formas en que el sesgo puede tener un impacto: el humano; el diseño de *software* y la forma en el que se usan los datos. En el mismo contexto, el desarrollo por parte de los Ayuntamientos del Reino Unido, en medio de una creciente crisis financiera, de los sistemas de "análisis predictivo" para identificar algorítmicamente a las familias que necesitan atención de servicios infantiles puede interferir en la intimidad de las personas y reforzar la estigmatización de determinados grupos de población.

Retos que supone el diseño de algoritmos éticos

Resulta complicado incluir principios éticos en la toma de decisiones algorítmicas dada su naturaleza. Antes de realizar ningún cambio, es importante responder a ciertas cuestiones: ¿Las presunciones éticas de los algoritmos deben ser transparentes y fáciles de identificar para los usuarios? ¿Qué pasa con los equipos de programación? ¿Son lo suficiente diversos? La gente a la que le afectarán estas cuestiones tendrán algún tipo de influencia sobre el sistema? Incluso si los algoritmos fuesen transparentes, ¿cómo podrían ser interpretados por accionistas que han recibido información destina sobre el tema? Pueden principios éticos como la justicia o el derecho a la privacidad ser codificados en el sistema y, si esto es así, cuáles deberían ser esos principios y quién

debería decidir sobre su elección? ¿Tiene nuestra sociedad un código moral universal que puede ser codificado?

Como respuesta a la oleada de noticias y artículos que tratan la naturaleza de la ética de los algoritmos, están empezando a aparecer certificados de garantía como el AI, Audit-AI de Pymetrics, Fairness Flow de Facebook y ORCAA. El reconocimiento de la necesidad de hacer operativo el juicio moral para el desarrollo de vehículos autónomos y de integrar agentes morales artificiales que puedan gestionar la complejidad de las nuevas tecnologías ha dado lugar a una serie de iniciativas de diseño de algoritmos basadas en diversas teorías éticas, como la de la National Science Foundation. La aparición de la inteligencia artificial y del aprendizaje avanzado de las máquinas puede llevar a que los coches automáticos estén encuadrados en un marco ético que podría establecer patrones de comportamiento clave. Mientras tanto, IBM ha desarrollado un nuevo conjunto de software de código abierto para ayudar a los desarrolladores a lidiar con los algoritmos de la caja negra y comprender cómo la inteligencia artificial que utilizan toma decisiones.

¿Qué implica el desarrollo de algoritmos para los líderes europeos?

Cuando nos referimos al impacto ético de los algoritmos en los mecanismos de toma de decisiones, aún no tenemos modelos ni procedimientos que regulen su funcionamiento.

Esto se debe, en parte, a la falta de normas existentes sobre las cuestiones en las que basar la certificación del algoritmo. Desarrollar principios éticos y códigos para los algoritmos implica identificar los principios, las normas, la asignación de roles y responsabilidades involucrados en la toma de decisiones.

La iniciativa global del IEEE para las consideraciones éticas en la inteligencia artificial y los sistemas autónomos, la Declaración de Toronto y el Fairness Flow de Facebook indican la necesidad de establecer objetivos de orientación social y puntos de referencia para el desarrollo de algoritmos. Sin embargo, dado que los algoritmos son de por sí amorales, la evaluación de su componente moral podría ser llevado a cabo por algorítmicas de impacto, de manera que las "auditorías algorítmicas" podrían acabar convirtiéndose en un requisito legal al implementar cualquier sistema de este tipo. Estas auditorías abordarían cuestiones éticas, como la legitimidad del uso de un sistema algorítmico de toma de decisiones en determinados contextos (por ejemplo, sentencias basadas en pruebas o armas letales), y serían elaboradas por comités éticos y organismos

de acreditación y certificación. Su objetivo debería consistir en evaluar los usos propuestos para la toma de decisiones algorítmicas en ámbitos de aplicación altamente sensibles y/o críticos para la seguridad e investigar los casos sospechosos de violaciones de derechos en el marco del mismo contexto tecnológico. Además, estos instrumentos podrían ayudar a los desarrolladores de sistemas y a los responsables de la toma de decisiones a revisar algunas de sus propias suposiciones de lo que realmente es un algoritmo.

Curiosamente, el requisito de que los responsables de la gestión de datos proporcionen a los interesados "información pertinente sobre la lógica implicada" en un proceso de toma de decisiones automatizado, introducido por el Reglamento general de protección de datos (RGP), puede allanar el camino para el desarrollo de una ética algorítmica práctica que aborde las virtudes, las consecuencias y la normativa. Aclarar los supuestos incorporados en el algoritmo o divulgar el código del sistema o la información sobre su lógica exige un examen cuidadoso de las normas pertinentes relativas a los derechos de propiedad intelectual que puedan limitar la accesibilidad. Para concluir, los responsables políticos de la UE tienen una oportunidad única de liderar el mundo en la regulación ética de la revolución digital, promoviendo el desarrollo de un marco ético general que rijan el diseño, la aplicación y el desarrollo de algoritmos. Estos deben permanecer bajo la supervisión y el control humanos y responder de manera satisfactoria a cuestiones como la parcialidad y otros posibles efectos.