Javier Gonzalez Vela

Licenciado en Ciencias Físicas por la Universidad Autónoma de Madrid.

Desde finales del siglo XX estamos asistiendo a una auténtica revolución tecnológica cuyo impacto en la sociedad es difícil de determinar en toda su extensión, pero cuyos primeros efectos son ya claramente visibles. La aparición de Internet, el despliegue masivo de los dispositivos móviles, la difusión de las redes sociales…, todo ello conlleva cambios de comportamiento profundos. Los avances en la capacidad de computación junto a las nuevas técnicas de tratamiento masivo de datos (que se conocen con el término técnico de "Big Data") proporcionan las herramientas necesarias para desarrollar sistemas que interactúan con nosotros de manera completamente nueva. Y los avances en robótica y en Inteligencia Artificial nos proporcionan nuevos dispositivos con los que tendremos que acostumbrarnos a convivir. Uno de los ejemplos más significativos, aunque no el único, lo constituyen los coches autónomos, capaces de conducir ellos solos sin requerir asistencia humana: son una muestra de que tanto la inteligencia artificial como los dispositivos que la implementan han alcanzado ya a un nivel de madurez tal que nos obliga a tomar decisiones concretas sobre cuestiones que normalmente abordaríamos de manera intuitiva. El comportamiento básico de las máquinas hay que programarlo, hay que escribirlo, y por tanto hay que especificarle a la máquina los criterios con los que debe comportarse ante determinadas situaciones. Es decir, hay que transcribir a algoritmos ejecutables por una máquina los criterios morales bajo los cuales tomamos determinadas decisiones. Y eso nos obliga a pensar cuáles son esos criterios morales, y por qué escogemos esos y no otros.

Para abordar estas cuestiones surgen iniciativas que, aprovechando las capacidades de Big Data, recaban información de un gran número de personas sobre sus preferencias y decisiones con la idea de utilizar las más extendidas como referencias morales.

Por Big Data entendemos el conjunto de técnicas que permiten hacer un procesado de cantidades ingentes de datos. Estos datos se obtienen ya fácilmente de la interacción con múltiples sistemas que nos rodean, y la capacidad de computación para procesarlos está ya disponible. Big Data te permite por tanto obtener patrones y correlaciones a partir de un número muy elevados de datos, y de ahí extraer conclusiones útiles.

Así nos encontramos, por ejemplo, con la encuesta de la *Moral Machine* elaborada conjuntamente por investigadores del CNRS francés (Centre National de la Recherche Scientifique), el MIT y las universidades de Harvard y de la Columbia Británica, o por el *Moral Sense Test* realizado por la Universidad de Harvard (y que ya han contestado más de 200.000 personas).

Centrándonos en la *Moral Machine*, posiblemente la más conocida puesto que han participado cerca de 2,5 millones de personas de 233 países distintos, se presenta vía web un formulario con múltiples escenarios en los que un coche autónomo está en situación de accidente inmediato por atropello, y debe decidir cómo actuar en cada caso, sabiendo que todas las decisiones implican salvar a unos seres y condenar a otros. Se trata de una variante del dilema del tranvía, en la que se presentan varios factores (salvar a personas frente a salvar animales, pasajeros frente a peatones, salvar más vidas frente a salvar menos, personas jóvenes frente a ancianas…, y así hasta nueve), y el participante tiene que elegir cuál es la opción mejor en cada caso. Utilizando el análisis de datos, se procesan todas las respuestas y se obtienen una serie de patrones de conducta, que se pueden agrupar por áreas geográficas, razas, sexos, o cualquier otro parámetro que se considere conveniente. Pero el sistema es aún más detallado, puesto que sitúa a cada persona en una escala comparativa respecto al resto de participantes, e incluso informa de cuáles han sido sus preferencias particulares que le han guiado en sus respuestas. Es indudable que te hace reflexionar sobre tus preferencias, pero… ¿podría un sistema así, al informarte de tus propias preferencias, llegar en un momento dado a condicionarlas? No es solo que te haga reflexionar sobre tus preferencias, sino que de alguna manera te las propone.

Sin entrar en un análisis detallado de los resultados (que se pueden consultar en el siguiente enlace), hay determinadas aspectos que merecen la pena destacarse: existen variaciones individuales importantes, especialmente cuando se considera el sexo y la religiosidad del participante; las preferencias están altamente correlacionadas con variaciones culturales y económicas entre países (considerando tres áreas de países claramente diferenciadas: occidentales, orientales y latinoamericanos); y además se notan diferencias notables entre culturas individualistas y culturas de colectividad. Pero aun así,

en el análisis de los resultados se observan tres criterios que se repiten de manera clara e inequívoca. La gente prefiere mayoritariamente:

1. salvar antes a las personas que a los animales.
2. salvar el mayor número posible de vidas.
3. salvar a personas jóvenes antes que a personas mayores.

Independiente de la valoración que nos merezcan estos criterios, cabe preguntarse si se trata de un método adecuado para establecer unas reglas morales. ¿El hecho de que una mayoría de personas los elijan los determina ya como criterios válidos? Y si fuera así, ¿en qué porcentaje se podría cifrar una mayoría válida?

Y no es que la *Moral Machine* pretenda establecer de esta manera los criterios (hasta la fecha, el único intento serio lo ha realizó en 2017 la "*German Ethics Commision on Automated and Connected Driving*"), pero sí intenta proporcionar una referencia válida para ello.

Tal vez si estuviésemos de acuerdo en la existencia de una moral universal válida sería más sencillo responder a estas preguntas, pero más allá de considerar que una elección moralmente correcta es la que maximiza el bien común, no parece haber un consenso claro sobre la existencia o no de una moral universal. Sirva como referencia la encuesta que sobre este tema *The Evolution Institute*(*) ha realizado entre 15 autores, entre los que se dan todas las opciones posibles: si, no, quizás…

Tal vez las respuesta a estas dudas se encuentre en las propias conclusiones de los impulsores de la *Moral Machine*, cuando dicen que «mientras que las preferencias éticas del público no deben ser necesariamente el primer árbitro de una política ética, la intención que tenga la gente de comprar coches autónomos y tolerarlos en las carreteras dependerá de la aceptación de las reglas éticas que se adopten», e igualmente cuando establecen que «antes de permitir que los coches tomen decisiones morales debemos iniciar un debate global para expresar nuestras opiniones a las empresas que diseñarán los algoritmos morales. La MM ha sido desarrollada para iniciar ese debate.»

Sin embargo, el método que utiliza el MIT en su *Moral Machine* parece ser bastante práctico: es un método eficiente en el sentido de que ofrece unas respuestas razonablemente rápidas y que pueden concitar los mayores acuerdos. Y lo que puede llegar

a ser más importante: puede acabar resultando una vía muy cómoda de tomar decisiones. Porque al final a las personas no les gusta tomar decisiones incómodas. Este es un efecto que se ve también claramente en el análisis de los resultados de la *Moral Machine*: en la variante del escenario en el que para salvar cinco vidas debes activar una palanca que mataría a una persona, el 90% de los encuestados preferiría no tener que activar la palanca.

Las decisiones son incómodas. Pero en el momento en que dispongamos de inteligencias artificiales, que además hayan sido capaces de aprender comportamientos de manera autónoma (y hay que decir que se está avanzando mucho en conseguir que las inteligencias artificiales aprendan rápidamente a partir de su propia experiencia), ¿no resultará muy reconfortante saber que alguien o algo podrá elegir por ti sin miedo a equivocarse? De una manera limpia y eficiente te pueden eximir de toda responsabilidad. ¿No sería ésta una situación que mucha gente encontraría deseable?

Al final podríamos estar sacrificando nuestra responsabilidad en aras de la eficiencia y la comodidad. ¿Nos llevaría esto a que una inteligencia artificial acabase sabiendo mejor que nosotros lo que nos conviene?

(*) *The Evolution Institute* es una organización sin ánimo de lucro con sede en San Antonio, Florida. Fundada en el año 2010 por David Sloan Wilson, biólogo evolutivo (área de la biología que estudia los cambios de los seres vivos a través del tiempo) y distinguido profesor de ciencias biológicas y antropología en la Universidad de Binghamton, EE.UU. *The Evolution Institute* busca aplicar la ciencia evolutiva a los problemas sociales. Y a través del proyecto Seshat hace un extenso de técnicas de tratamiento masivo de datos para comprobar hipótesis científicas. (fuente: Wikipedia).

Referencias:

Anónimo (s.f.) Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad. En *Power Data*. Accesible el 25/11/2017 en https://www.powerdata.es/big-data

Miranda-Saavedra, Diego (2018). De cuando la filosofía rescató a la inteligencia artificial. En *Retina. El País*. Accesible el 25/11/2018 en (https://retina.elpais.com/retina/2018/10/30/tendencias/1540906405_872503.html#?id_externo_nwl=newsletter_retina20181108

Malo, Pablo (2018) ¿Existe una moral universal? En Evolución y Neurociencia. Accesible el

25/11/2108 en
https://evolucionyneurociencias.blogspot.com/2018/06/existe-una-moral-universal.html

Plaza López, José Ángel. (2018) Bioética. El peligro de vivir en una sociedad digital diseñada para manipular. En *Retina. El País.* Accesible el 25/11/2017 en
https://retina.elpais.com/retina/2018/10/02/tendencias/1538462550_862834.html

Torre, Diego de la y Martín-Peñasco, Jaime (2018) ¿Qué es el aprendizaje profundo? Así es el método de obtención de conocimientos de la Inteligencia Artificial. En *Blogthinkbig*. Accesible el 25/11/2018 eb
https://blogthinkbig.com/aprendizaje-profundo-inteligencia-artificial

Si desea citar esta página

Gonzalez Vela, Javier (2018). Implicaciones morales del estudio Moral Machine.
En *Niaia,* consultado el 28/11/2018
en https://www.niaia.es/implicaciones-morales-del-estudio-moral-machine/