



## Ética y transparencia para la detección de sesgos algorítmicos de género

Lucía Benítez Eyzaguirre<sup>1</sup>

Recibido: 20 de septiembre de 2018 / Aceptado: 28 de enero de 2019

**Resumen.** La creciente importancia de los algoritmos pone en evidencia la discriminación que se registra, especialmente en género y minorías, así como la necesidad de transparencia en la aplicación de estas fórmulas frente a la opacidad de las corporaciones. A pesar de estos sesgos, en los algoritmos se apoya la toma de decisiones de casi todos los campos del conocimiento y de las actividades sociales, políticas y económicas a causa de una confianza casi ciega en los procesamientos informáticos, de un imaginario tecnológico sobre su capacidad para eliminar el error y el sesgo. El Efecto Manipulador del Motor de Búsqueda o Search Engine Manipulation Effect (SEME) (Epstein y Robertson, 2015) muestra efectos muy claros sobre el comportamiento del voto. También Caliskan y Bryson (2017) han detectado la reproducción de sesgos de género y étnicos, a partir de datos ya sesgados que conducen a desviaciones estadísticas muy importantes en el Big Data.

**Palabras clave:** Transparencia; algoritmos; conocimiento abierto; género.

### [en] Ethics and transparency for detection of gender bias in algorithms

**Abstract.** The algorithms' growing importance shows the discrimination registered, especially on gender and minority groups, besides the need of transparency in the application of these formulas against the corporations' opacity. Despite these biases, the making decision on almost all the knowledge fields, as well as the social, political and economic activities, leans on algorithms because of the blind trust in computer processing and the technological imaginary about their ability to eliminate the error and the bias. The Search Engine Manipulation Effect (SEME) (Epstein y Robertson, 2015) shows very clear effects on voting behavior. Caliskan y Bryson (2017) have also detected the reproduction of gender and ethnic biases when working on already biased data, which lead to a very important statistical deviations in Big.

**Keywords:** Transparency; algorithms; open knowledge; gender.

**Sumario.** 1. El poder de los algoritmos. 2. Tipos de algoritmos. 3. Algoritmos sociales. 4. La industria de los datos. 5. La economía del dato y el trabajo inmaterial. 6. Los sesgos y la perspectiva feminista. 7. El Efecto Manipulador del Motor de Búsqueda. 8. Efectos discriminatorios de género. 9. Modelo opaco y poco democrático. 10. Transparencia y ética en los algoritmos. 11. Soluciones humanas: educación y verificación de datos. 12. Conclusiones. 13. Referencias bibliográficas.

<sup>1</sup> Universidad de Cádiz (España)  
E-mail: lucia.benitez@uca.es

**Cómo citar:** Benítez Eyzaguirre, Lucía (2019): "Ética y transparencia para la detección de sesgos algorítmicos de género". *Estudios sobre el Mensaje Periodístico* 25(3), 1307-1320.

## 1. El poder de los algoritmos

El algoritmo como procedimiento sistemático y efectivo para resolver un problema se compone de un número concreto de instrucciones que deben estar bien definidas para operar sobre un tipo concreto de datos, a través de un número finito de pasos, que aporta una solución —generalmente matemática— a todos los casos analizados. La solidez de la formulación de un algoritmo dependerá de que sea sistemático y consistente, pero también de que sea no ambiguo, es decir que no deje abierta la solución a interpretaciones. La misión de un algoritmo no es el cálculo de una solución exacta o precisa, sobre todo porque como objetivo se convierte en inalcanzable en muchos casos; su misión está más próxima a facilitar una respuesta satisfactoria o válida, es decir, útil más que en su neutralidad o imparcialidad.

Eli Pariser (2017) ha denominado la “burbuja de filtros” al efecto por el algoritmo selecciona las informaciones que prefiere el usuario, en función de las búsquedas anteriores, la geolocalización, o las preferencias en cualquier tipo de selecciones. La burbuja presenta una web empequeñecida con los mismos contenidos que coinciden con los propios puntos de vista, reforzando una ideología previa y al margen de un pensamiento crítico. Pariser propone que los algoritmos de estas plataformas funcionen de modo compensatorio para respetar la pluralidad de visiones en el acceso a la información. Con los años, el fenómeno se ha vuelto más extremo ya que también se consideran las noticias y contenidos escogidos por personas cercanas como amigos o familiares, acabando así con la idea de una web que ofrecía información de forma abierta, descentralizada y no jerárquica.

## 2. Tipos de algoritmos

Las capacidades de los algoritmos a la hora de resolver cuestiones conducen a una visión diferentes: no son ya herramientas, sino parte del proceso que facilita un diseño automatizado, es decir, una actualidad en función de del procesamiento de los datos porque permite el diseño actualizado de muchos tipos de procesos:

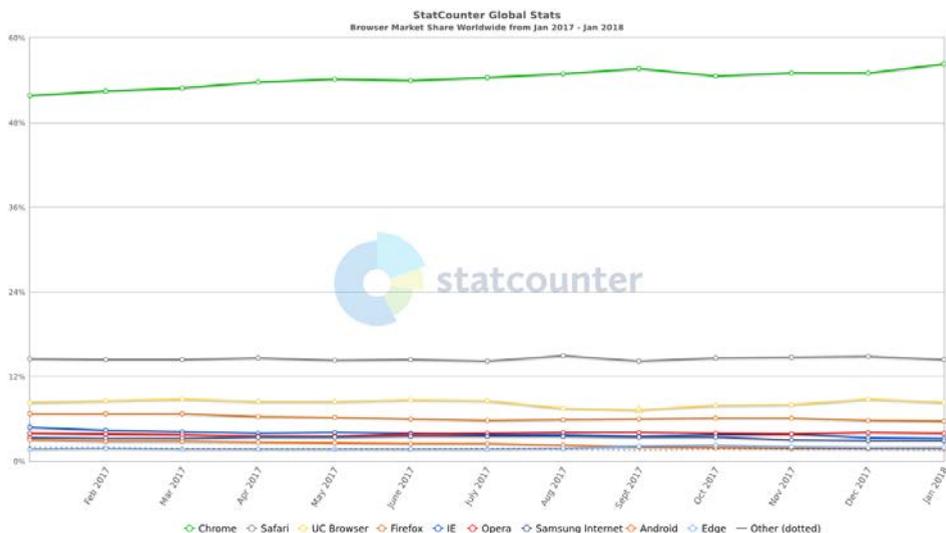
“Algorithms are no longer seen as tools to accomplish a task: in digital architecture, they are the constructive material or abstract stuff that enables the automated design of buildings, infrastructures, and objects. Algorithms are thus actualities, defined by an automated prehension of data in the computational processing of probabilities” (Parisi, 2013: xii)

Los algoritmos buscan la mejor solución a un problema por diferentes vías, y en función de ello serán: Algoritmos deterministas, que condicionan de forma única el siguiente paso en su formulación, por ello produce una respuesta que se puede conocer en función de los datos de entrada, el algoritmo determinista siempre encuentra la misma solución. A este tipo de algoritmos no se les permite que calcule una solución incorrecta ni que no termine de realizar sus operaciones. Por

otra parte, los algoritmos no deterministas ofrecen muchos resultados posibles, y por tanto no ofrecen una solución única, ni se puede conocer a priori cuál será el resultado que ofrezca por lo que el análisis de estos algoritmos probabilistas es, muy a menudo, muy difícil. Los resultados incorrectos de estos algoritmos pueden ser útiles sobre todo si la respuesta que producen es la solución aproximada de un problema especialmente difícil.

Los lenguajes de programación tan conocidos como Pascal, Basic o C++ utilizan en su fase de experimentación datasets de textos en lenguaje humano en su mayoría en inglés. Las máquinas de aprendizaje automático incorporan el lenguaje sexista y racista a través de los datos que le sirven para representar hechos, para mostrar la cultura, para aportar soluciones (Caliskan et al., 2017). El análisis de Caliskan, Bryson y Narayanan (2017) sobre dos millones de palabras utilizadas en el aprendizaje automático y su asociación con diferentes términos los llevó a la conclusión de que se registraba una reproducción de los sesgos humanos. Estadísticamente es más probable que las palabras femeninas o los nombres propios de mujer se relacionen con el mundo privado, las tareas domésticas, lo artístico o emocional, mientras que lo masculino se asocia con lo profesional, lo científico y el conocimiento. En el procesamiento de grandes datos se dan estereotipos muy significativos en los que basa su aprendizaje la AI. Aún lo es más si se tiene en cuenta que la ideología de las personas que deciden el problema a resolver con el algoritmo, sus objetivos y requisitos, también genera sesgos en el propio proceso de formulación. El riesgo, por tanto, está en los datos y en la falta de supervisión. Mucho más si, como en muchos campos de la actividad humana, son personas occidentales de género masculino cuya visión del mundo se traslada automáticamente al diseño de las estructuras y reglas del sistema a veces de forma inconsciente, pero también como fruto de la experiencia personal.

Gráfico 1. StatCounter Global Stats. Fuente: StatCounter



Profundizar en el funcionamiento y los efectos de los algoritmos choca con la lógica mercantilista y de apropiación con que han desarrollado el modelo las grandes empresas tecnológicas. En avanzadilla, Google desarrolló su algoritmo PageRank otorgando a las páginas web un valor que se basaba en los vínculos de otras páginas relacionadas y en la clasificación de estas últimas. Así fue como el buscador consiguió su posición de predominio en el mercado, que todavía mantiene, como muestran los resultados de páginas de analítica de tráfico como StatCounter<sup>2</sup>. La preeminencia del buscador obliga a las webs a lograr indexarse y buscar su posicionamiento, para aparecer entre los primeros resultados.

Google ofrece resultados sobre un rastreo en más de 60 de trillones de páginas que actualmente tiene la web, y decide, con su algoritmo, cuáles serán las páginas que aparecerán en nuestra lista de resultados y su orden. El ranking general está organizado por unas doscientas mil variables, mientras la personalización se afina con la asistencia de otros algoritmos que indexan informaciones relativas al usuario, tales como como la localización, tipo de buscador, el historial de búsquedas o el idioma. La empresa explica la automatización del procedimiento en una web<sup>3</sup>, a la vez que advierte de los cambios que realiza en algoritmo en función de propuestas de sus equipos de trabajo<sup>4</sup>.

Desde 2013, Google utiliza un algoritmo semántico, Hummingbird, diseñado para el análisis del lenguaje natural o búsqueda conversacional. Con este nuevo algoritmo, la compañía actualizó los criterios incorporando la relevancia social, los factores personales, la calidad del contenido y los enlaces, así como la elección personal del usuario a la hora de dirigirse a uno de los resultados. En este sentido, Sweeney (2013) cuestiona cómo este modelo expone a las audiencias a prejuicios raciales y de género, a la vista de cómo se publica la publicidad contextual.

Dos años después Google implementó el componente RankBrain, para analizar un quince por ciento de las búsquedas, cuando incluyen términos coloquiales, frases complejas o neologismos, y así anticiparse a la intención del usuario. Entonces también incluyó la herramienta de código abierto TensorFlow, que aplica el modelo matemático Word2Vec, con el que se inicia el aprendizaje de conceptos a partir de las lecturas de artículos y sin supervisión humana. Usa representaciones distribuidas del texto para detectar similitudes entre conceptos, de forma que se hace útil para representar y extraer conocimiento, la traducción automática, la formulación de preguntas y respuestas (Mikolov et al., 2013).

### 3. Algoritmos sociales

El fundamento de este tipo de algoritmos es la comunicación estigmérgica, que permite la colaboración entre individuos a través del medio físico sin necesidad de una planificación o un poder central. Se inspira en el comportamiento de insectos, como las hormigas, que se comunican a través de su rastro con feromonas, la temperatura o bien los propios objetos que desplazan, una disciplina que se ha

---

<sup>2</sup> <https://statcounter.com/>

<sup>3</sup> <https://www.google.com/search/howsearchworks/>

<sup>4</sup> En junio de 2017, la Comisión Europea sancionó a Google con 2420 millones de euros por redirigir las búsquedas hacia su portal comercial Google Shopping.

convertido en una técnica probabilística que recibe el nombre de Ant Colony Optimization (ACO), soluciona problemas de computación al localizar los mejores caminos o rutas en grafos. Imitando el hallazgo de una solución local a un problema complejo, cada vez que compartimos, buscamos, aprobamos, compramos o twitteamos algo, se registra en datos que se procesan, sobre los que reciben información otros usuarios y son consumidos por la colectividad. Al gestionar la información, los algoritmos reconstruyen las relaciones, la organización, los gustos y las identidades.

La conexión entre usuarios en las redes se establece también en relación con prácticas sociales como la puntuación, la clasificación y la recomendación (Brinton y Chiang, 2017). La puntuación combina el puntaje de la relevancia y de la importancia e influye en los resultados, pero también según la reputación de la persona que comenta, la disparidad de las puntuaciones y otros elementos, como el historial de navegación, logrando una presentación personalizada de los resultados. La recomendación, que se utiliza sobre todo para productos culturales como películas y libros, llega a ser predictiva por comparación con usuarios parecidos y la puntuación obtenida por el producto.

#### 4. La industria de los datos

Las corporaciones tecnológicas defienden que los filtros y la personalización son esenciales para dar eficacia a los servicios y ordenar las preferencias. Bajo la premisa de que más datos producen mejores algoritmos y mejores servicios, el usuario se ha convertido en el producto con el que se comercia. El estudio elaborado por IDC “Worldwide Big Data Technology and Services, 2012–2015” estima el valor de la economía del dato en Europa en los 272.000 millones de euros en 2015, lo que representa el 1,87% del PIB de los países miembros y podía ascender al 4,7% en 2020. Estimado desde el valor de productos y servicios relacionados con los datos se situaría alrededor del 25% de 2016 a 2021.

Mientras las decisiones basadas en los análisis de datos proporcionados por algoritmos generen más beneficios o más producción a las empresas será difícil que cambien las cosas. McAfee y Brynjolfsson (2012) estimaron que las empresas que utilizan decisiones basadas en algoritmos aumentaban su rentabilidad en un 6% y eran un 4% más productivas. Dos tercios de las operaciones realizadas en bolsa se realizan a través de algoritmos, ya que la velocidad es siempre una ventaja sobre la competencia. De hecho, cinco de las mayores empresas de software -Facebook, Apple, Amazon, Netflix y Google- capitalizaban el 10% del índice S&P 500<sup>5</sup> en 2016.

La automatización de las decisiones está alcanzando todos los sectores de la economía para prevenir el riesgo de que las empresas puedan resultar obsoletas. La ambición es reprogramar el mundo acorde con este sistema, aunque las consecuencias se registren en todos los ámbitos, desde los negocios hasta la

---

<sup>5</sup> El índice Standard & Poor's 500 (Standard & Poor's 500 Index) se considera el más representativo del mercado bursátil ya que se basa en la capitalización de las mayores empresas en la bolsa de Nueva York y en la Nasdaq, la segunda mayor automatizada. Se considera un indicador importante por su seguimiento y por las tendencias que marca de cara al valor de las acciones en Estados Unidos.

política. Estos cambios se fraguan al margen de la ciudadanía y de la sociedad libre que, si no participa, no podrá controlar el proceso. Se procesa información sobre cómo pensamos, qué hacemos o cuáles son nuestras emociones para pasar a convertirse en datos, la materia prima de la economía digital que, bajo la lógica numérica, se transforma en un nuevo sistema de producción. Desde la perspectiva feminista, se capitaliza así el trabajo afectivo y de conocimiento de nuestras interacciones sociales como un insumo de importancia para las últimas formas del capitalismo.

## **5. La economía del dato y el trabajo inmaterial**

Como materia prima, los datos no se pueden concentrar en pocas manos, en las cinco empresas más cotizadas que ocupan en conjunto en el mercado una posición dominante, que han logrado facilitando servicios de gran interés para determinados sectores y para la propia ciudadanía, con su política de gratuidad en lo que al pago en moneda se refiere. En la cadena de valor de la economía, el mundo digital ha convertido a los demás sectores y a los datos proporcionados por éstos en su materia prima, de la misma forma que el sector servicios hizo con la economía primaria y secundaria. Así han logrado, bajo una lógica de aparente desintermediación, convertirse en los grandes intermediarios entre mercados, sectores y actividades.

Los datos están, por tanto, cargados de poder y la potencia aumenta porque la lógica del mundo en red es de atracción: los datos llaman a más datos, a la concentración. La lógica capitalista transforma la automatización y la participación en nuevos activos productivos. Al igual que los logros de las luchas sindicales del siglo XX —más salario y más tiempo libre para la clase trabajadora— han servido para sentar los pilares de la sociedad de consumo —más tiempo libre para ver la televisión y su publicidad, más dinero para comprar en los centros comerciales— ahora el excedente de tiempo que se libera y también la energía se introducen de nuevo en el ciclo productivo y de acumulación de riqueza.

La transformación del algoritmo en un software social precisa también de la adaptación a los cuerpos y a los comportamientos humanos, que mejoraría su existencia autónoma. Desde el feminismo hay que indagar, más allá de la monetización y de los sistemas de control que la acompañan, el tiempo y la energía que la automatización consume de cara al común para afrontar la racionalidad política ante los constantes cambios que proponen los algoritmos.

## **6. Los sesgos y la perspectiva feminista**

Desde el feminismo los datos personales son una forma de política en la que debería entenderse que la ciudadanía tiene intereses legítimos para exigir libertad e igualdad personal y política. Así como también entenderse que la privacidad es “una habitación propia”, como la de Virginia Wolf, en la que es válida la

formulación del número 13 de los Principios Feministas de Internet de la Asociación para el Progreso de las Comunicaciones (APC)<sup>6</sup>:

“Apoyamos el derecho a la privacidad y al control total sobre los datos y la información personal en línea en todos sus niveles. Rechazamos las prácticas de los estados y las empresas privadas que utilizan los datos con fines de lucro y para manipular el comportamiento en línea. La vigilancia es la herramienta histórica del patriarcado, utilizada para controlar y restringir los cuerpos, las voces y el activismo de las mujeres. Lo mismo vale para las prácticas de vigilancia por parte de individuos, sector privado, actores estatales y no estatales”.

El feminismo no entra en contradicción si apoya la explotación de los datos para el bien común, lo importante es cuestionar las relaciones de poder y los derechos humanos: De quién se obtiene la información, con qué fin, qué algoritmos o cálculos se utilizan, es decir, las cuestiones relacionadas con el consentimiento, el acceso, la propiedad y, por tanto, la agencia en la que hay que incorporar —y compartir— los conocimientos sobre la privacidad, la navegación anónima y la encriptación. La decisión sobre cuáles son los datos públicos es central en el sistema, una formulación política que debería estar respaldada por un contexto jurídico institucional. Incluso para el caso en que haya inclusión de los datos, hay mucho por decidir sobre el tema porque los datos inclusivos no son suficientes para la igualdad, el empoderamiento y las minorías, ya que la gestión de los datos es esencial en el contexto actual para la consecución de derechos sociales, económicos, políticos y culturales.

Los movimientos feministas de carácter local o de pequeñas comunidades son una fuerza disruptiva ante los modos de gestionar los datos por las corporaciones e instituciones, ya que presentan otros modos de gobernanza, otros imaginarios para la justicia de género y la democracia. Es un desafío y una forma de resistencia ante las políticas de datos optimizadas con criterios ajenos. Estas posiciones, al margen de la disidencia y resistencia, son imprescindibles para la mejora del sistema, ya que las personas que quedan fuera de los resultados, o que no obtienen la suficiente representatividad, resultan ser las mismas víctimas de homofobia, sexismo o racismo.

El origen del problema se encuentra en los *datasets* más utilizados para los *benchmarks* que miden el buen funcionamiento de los algoritmos de reconocimiento facial que son mayoritariamente masculinos (77%) y de origen caucásico (entre el 79 y el 86% de rostros blancos). Por ello clasifican mejor a los hombres blancos que a las mujeres de otras etnias (Buolamwini y Gebru, 2018). La investigación de Datta et al. (2015) mostró cómo, a partir del desarrollo de la herramienta AdFisher, el buscador Google segmenta la publicidad en función del comportamiento de los usuarios, se ofertan los puestos de trabajo de mayor categoría y salario a hombres blancos. Edelman y Luca (2014) mostraron la discriminación racial, de género, aspecto y edad que se registra en la plataforma Airbnb, que favorece operaciones económicas más rentables de alquiler a los

---

<sup>6</sup> <https://www.apc.org/es/pubs/principios-feministas-para-internet-version-2>

hombres blancos, no sólo a través del algoritmo sino de las puntuaciones de los usuarios sobre los perfiles personales.

## **7. El Efecto Manipulador del Motor de Búsqueda**

Epstein y Robertson (2015) han investigado los resultados de búsqueda para comparar si los resultados de diferentes usuarios benefician a candidatos electorales, a través de experimentos controlados en los que entre el 50 y el 60 % de los participantes cambiaron su voto por el candidato que se situaba en el primer resultado de un motor de búsqueda, y entre el 99,5 y el 75 % de estas personas no tuvieron consciencia de que los resultados estaban manipulados. Epstein y Robertson (2015) defienden que Google tiene posibilidad de cambiar el resultado del 25 % de las elecciones nacionales de diferentes países del mundo sin que nadie advierta si se produce una manipulación en los resultados, como de hecho están convencidos de que ya se ha producido.

Así llegaron a la definición del Efecto Manipulador del Motor de Búsqueda o Search Engine Manipulation Effect (SEME) (Epstein y Robertson, 2015) sobre el que aseguran que es uno de los mayores efectos sobre el comportamiento humano que se haya sistematizado. Por una parte, este efecto no sólo pasa desapercibido, sino que los resultados de una búsqueda se interpretan como excelentes. El rastreo se agrava debido a que los resultados de búsqueda son fugaces porque no dejan huella y pueden ser modificados y también negados por la empresa, ya que no existe ninguna transparencia sobre los criterios del algoritmo.

## **8. Efectos discriminatorios de género**

Frente a la resolución de la ONU (2016) en defensa de los derechos humanos en Internet, en la que se insiste en la necesidad de igualdad, el desarrollo de los algoritmos perpetúa una situación discriminatoria para las mujeres y las minorías. Los algoritmos se reprograman con la información que obtienen y producen reglas externas, es decir, que ya no son una garantía del pretendido orden y control que sugiere la automatización. Así lo entiende Luciana Parisi (2013: IX) cuando mantiene que los algoritmos, considerados como modelos racionales de gobernanza, no son confiables para la programación de tareas, ni para reproducir reglas ni para analizar comportamientos: “Randomness has become the condition of programming culture”.

Todo ello, y por contradictorio que parezca, coincide con muchas iniciativas basadas en el aprovechamiento del Big Data que también generan un enorme potencial para reducir la brecha de género, para potenciar la consecución de derechos humanos, para mejorar el desarrollo y la vida de las personas. De hecho, la hipótesis sobre la que se apoya el uso generalizado de datos, la obtención de mejores y más datos es que éstos son una vía para alcanzar los Objetivos del Desarrollo Sostenible, a través de un proceso de toma de decisiones basadas en esa información. Iniciativas como United Nations Global Pulse u Open Algorithms Project trabajan en ese sentido, pero sólo aportará resultados conforme su diseño y

las soluciones que aporten incluyan la inclusión y representación de mujeres y niñas, de minorías, porque son estos colectivos los que tienen potencial para generar cambios en positivo.

## **9. Modelo opaco y poco democrático**

El informe de Pew Research Center (Rainie & Anderson, 2017) considera imprescindible, a la vista de la creciente importancia e impacto de los algoritmos, la formación sobre su diseño y funcionamiento, así como sobre sus implicaciones no sólo por la cantidad de operaciones que realizan a diario sino por la percepción que se produce alrededor de sus resultados. Para los usuarios será imposible detectar los sesgos o conocer si su clasificación ha sido justa o no. Para la actual alfabetización sería imprescindible conocer el papel de los algoritmos, el impacto de las valoraciones sociales —a través de mecanismos como la puntuación, de la clasificación y de la recomendación—, porque tienen una gran incidencia en los resultados del análisis de la información que realizan los buscadores.

Junto a ello, el informe de Pew Research Center (Rainie & Anderson, 2017) llama la atención sobre el conocido como ‘efecto burbuja’, que reduce la diversidad en el acceso a la información, efecto que se acentúa conforme los algoritmos van teniendo más información sobre los usuarios. El efecto se conoce como “sesgo de confirmación” y produce la sensación de que nuestro pensamiento es dominante por la ausencia de puntos de vista diferentes, algo que limita e incluso destruye el debate social. La falta de transparencia de los buscadores oculta información sobre la procedencia y el tratamiento de la información y, por supuesto, del algoritmo, considerado como un secreto industrial (Cadon, 2017: 39). Mientras se ha generalizado el uso de herramientas, cada vez más capaces y de menor coste, el acceso a los datos es privativo sin que se garantice la “governabilidad algorítmica” (Rouvroy y Berns, 2013) como el modo de regular las conductas y las repercusiones que tiene sobre la obediencia de las personas. La gobernabilidad algorítmica nos permitiría la transparencia, la supervisión colectiva de usuarios y empresas, la responsabilidad ante el uso de la información, o sea, la prevención del abuso y del control, la socialización de los beneficios para beneficio del Procomún.

El código es político (Cadon, 2017) y se introduce en nuestra realidad mientras no hay estrategias civiles para hacerle frente, ya que ni siquiera se contempla su regulación, la adopción de un estándar que devuelva la importancia de su impacto. La autonomía es imposible ante un sistema opaco en el que los algoritmos marcan las elecciones, como ocurre con la publicidad, cuya proximidad a nuestros gustos e incluso nuestros intereses es tanta que nos trata como consumidores y no como ciudadanos.

## **10. Transparencia y ética en los algoritmos**

Las características que debe reunir un algoritmo para que garantice la transparencia y calidad de los datos que facilita pasan por ser abierto y nutrirse de datos también

abiertos (Open Data), que haya sido creado con código de fuente abierto para que pueda ser auditable, y que sea alterable, para que pueda responder a las reclamaciones fruto de los cambios en su funcionamiento. Este algoritmo debe ser transparente, o sea, permitir auditorías sistemáticas sobre sus cambios u operaciones, incluso sometido a controles ciudadanos (Cadon, 2017: 41). Por último, debe ser leal y justo, evitar la discriminación. Las exigencias no terminan ahí: si se trata de un algoritmo de uso online, hay que tener también en cuenta las API (Application Public Programming Interface) públicas porque condicionan tanto la recolecta de datos como aplicar técnicas comerciales. Algunas iniciativas contemplan los estándares abiertos, como la que impulsa Tim Berners-Lee con el nombre de SOLID7 para construir una red global de datos abiertos –linked open data cloud-, a partir de la anterior ‘linked (open) data’. En la misma línea y en el MIT, Openpds/SA8 propone la privacidad y el control sobre los ‘personal data stores’, pero sin el mismo grado de apertura<sup>9</sup> ya que no está comprometido con el estándar de ‘linked data’.

Para responder a la situación se ha desarrollado la “Ley de la república digital”, que entró en vigor en Francia en 2016, a partir de un debate sobre la transparencia. También, la Association for Computing Machinery US Public Policy Council (USACM, 2017) ha propuesto siete principios para el control de los algoritmos: la concienciación sobre los posibles sesgos presentes en sus procesos; la impugnación y compensación por el uso de los algoritmos; la responsabilidad de las decisiones adoptadas a través de su uso; la transparencia sobre los procesos y decisiones, especialmente de cara al sector público; el control de la captación y procedencia de los datos con su revisión pública (aunque también figura una forma restringida a expertos); la posibilidad de auditar los resultados, y, por último, la validación y la prueba para conocer el impacto discriminatorio de sus fórmulas<sup>10</sup>. Sin embargo, la transparencia total no estaría garantizada, dada la necesidad de proteger tanto datos personales como las patentes industriales.

La Unión Estadounidense por las Libertades Civiles (American Civil Liberties Union, ACLU), dentro de sus objetivos relacionados con la defensa de los derechos individuales y las libertades de la ciudadanía, ha lanzado una iniciativa relacionada con los sesgos algorítmicos: el proyecto AI Now<sup>11</sup>, con el que investigar las consecuencias sociales, humanas y económicas de la Inteligencia Artificial. En su último informe, AI Now 2017 Report, lanza recomendaciones para garantizar un desarrollo de la Inteligencia artificial libre de sesgos, para lo cual defiende la necesidad de la verificación humana por parte de equipos de trabajo en los que se respete la diversidad.

---

<sup>7</sup> <https://solid.mit.edu/>

<sup>8</sup> <http://openpds.media.mit.edu/>

<sup>9</sup> Otras iniciativas de control de los datos privados, más agnósticas con respecto a la tecnología a usar, son: <https://thegooddata.org/> (información de nuestra navegación por Internet), <https://midata.coop/> (información de salud), o <http://www.greenbuttondata.org/> (información de consumo de energía)

<sup>10</sup> IEEE ha creado un grupo de trabajo y una propuesta de estándar para tratar el tema de la ética, el sesgo y la transparencia en los algoritmos (Koene, 2017)

<sup>11</sup> <https://ainowinstitute.org/>

## 11. Soluciones humanas: educación y verificación de datos

El antropólogo Pierre Lévy (2013) lleva años trabajando en el estudio del ‘medio algorítmico’ como la estructura automatizada de manipulación que será el fondo de la memoria común. Lévy busca un sistema simbólico que responda al cambio actual y sea capaz de explotar el pensamiento computacional, la memoria y la ubicuidad de Internet. Con pensamiento crítico quiere llevar la transparencia a la información, con un cambio al procesamiento de los algoritmos que relacionan los datos con modelos lógicos y de álgebra -el actual sistema en que se basa el Big Data-, para enlazar de forma general y universal los ecosistemas de ideas de las personas, o inteligencia colectiva que construya una memoria digital participativa, común a todas las personas. Para ello ha desarrollado el Information Economy Meta Language (IEML), un metalenguaje que hace computable la dimensión semántica, a través de un sistema sintáctico de coordenadas que puede transformar la información en conocimiento. La categorización y la evaluación de los datos partirán del desarrollo libre del diálogo, de la intersubjetividad y de la pragmática, de una forma transparente e interoperable. El diccionario de IEML, que pretende ser colaborativo, garantizaría la traducción a diferentes lenguajes naturales, a partir de las redes de relación semántica. El proyecto, de una gran complejidad, fomenta la interoperabilidad semántica y genera ecosistemas de ideas a partir del uso de las comunidades en línea.

Otras iniciativas tratan de atajar el problema de los sesgos. El proyecto europeo FA\*IR, con participación de la Universidad Pompeu Fabra y Eurecat, ha investigado la discriminación según género, procedencia o apariencia que lleva a invisibilizar a determinados colectivos en la búsqueda de empleo, un problema común a otros algoritmos utilizados por plataformas de economía colaborativa y redes sociales, y especialmente a las mujeres, ya que entre los primeros puestos de los resultados sólo aparecen hombres. El equipo de investigación ya ha desarrollado un algoritmo de búsqueda que invalida el sesgo étnico, de género o de edad, sin afectar al ranking obtenido que financiará el Data Transparency Lab.

Otras alternativas se han centrado en indicadores que cuantifican y definen los sesgos, directos o indirectos, cuando se insertan en el software de un buscador o en otras utilidades basadas en el Big Data, con un modelo en la práctica (Bolukbasi et al., 2016) que se puede utilizar sin modificar las propiedades útiles del sistema. Mathew

Fuller (2003) reclama el análisis crítico y humano de los algoritmos con métodos como la transparencia de su funcionamiento y condiciones de verdad, la conexión al software social, el enriquecimiento de conexiones para la mejora, y el impulso a su uso. La Organización para la Cooperación y el Desarrollo Económico (OCDE) incorporará el análisis crítico de la Información digital al próximo informe PISA sobre el rendimiento de los estudiantes de secundaria, en una comparativa de carácter internacional. Otras organizaciones como el Future of Life Institute trabaja para reducir los riesgos de la Inteligencia Artificial y ha elaborado una serie de

principios éticos<sup>12</sup> que definen, en conjunto, las necesidades del desarrollo de instrumentos e información de acuerdo con la seguridad, la transparencia, la responsabilidad, el alineamiento con valores humanos, la privacidad, la libertad, el control humano y el beneficio y la prosperidad comunes.

## 12. Conclusiones

Dado que el acceso a la información sobre el comportamiento social es la clave del control de los mercados —y puede llegar a actuar como un monopolio—, hay que analizar los elementos discriminatorios y de sesgo para valorar su impacto. De una parte, sobre la cultura, ya que la datificación, con sus algoritmos y sesgos, produce efectos de homogeneización cultural y limita la diversidad. De otra, porque la modelización basada en la conectividad de datos tiene consecuencias destructivas sociales y económicas y, por tanto, impactos sobre todo en las mujeres. El desarrollo algorítmico y el crecimiento de los sesgos en sus resultados conducen a una serie de recomendaciones y mejoras para limitar su impacto:

1. La necesidad de supervisión humana de la verificación de los datos y de los resultados.

2. La diversidad en la composición de los equipos de revisión

3. La comprobación y eliminación de los sesgos en la compra de datos, así como la exigencia en los contratos con proveedores de cláusulas contra los sesgos.

4. La adopción de técnicas que eviten la discriminación, potenciando un sistema de clasificación de la información legítima para dar mayor calidad a los datos (Hajian y Domingo-Ferrer, 2013).

5. La transparencia en las políticas de privacidad y en el consentimiento de los usuarios sobre el uso de datos.

6. El testeo de los algoritmos contra los sesgos, con ‘desaprendizaje’ de relaciones causales y de serendipia que refuerzan la presencia de comportamientos anteriores.

7. La aplicación del Análisis de Componentes Independientes (ICA) para revelar factores ocultos en las variables, que evalúen el impacto social, cultural y político de los resultados que ofrecen los algoritmos.

8. La investigación sobre las soluciones facilitadas por algoritmos para conocer su calidad y sus impactos, para adoptar decisiones de carácter ético.

Además de estas recomendaciones, todo evidencia la necesidad de implicarse, aprender y comprender los efectos de los algoritmos, así como de otras transformaciones de la digitalización. En caso contrario, estaríamos en manos de quienes los formulan y de la propia tecnología, quienes en principio nunca estarán interesados en las libertades ciudadanas.

---

<sup>12</sup> <https://futureoflife.org/ai-principles/>

### 13. Referencias bibliográficas

- Association for Computing Machinery US Public Policy Council [USACM] (2017): "Statement on Algorithmic Transparency and Accountability" (January, 12). En: [http://www.acm.org/binaries/content/assets/public-policy/2017\\_usacm\\_statement\\_algorithms.pdf](http://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf)
- Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James Y.; Saligrama, Venkatesh; and Kalai, Adam T. (2016): "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings". *Proceedings of Advances in Neural Information Processing Systems*, 29 (NIPS 2016), 4349-4357.
- Buolamwini, Joy and Gebru, Timnit (2018): "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". *Proceedings of Machine Learning Research*, 81, 1–15. Conference on Fairness, Accountability, and Transparency.
- Brinton, Christopher G., and Chiang, Mung (2017): *The Power of Networks: Six principles that connect our lives*. Princeton and Oxford: Princeton University Press.
- Cadon, Benjamin (2017): "El código es político, los algoritmos son armas matemáticas de destrucción". *Soberanía tecnológica*, 2, 31-47.
- Caliskan, Aylin; Bryson, Joana; and Narayanan, Arvind (2017): "Semantics derived automatically from language corpora contain human-like biases". *Science*, 356 (6334), 183-186. Doi: 10.1126/science.aal4230
- Datta, Amit; Tschantz, Michael Carl; and Datta, Anupam (2015): "Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination". *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2015 (1). Doi: <https://doi.org/10.1515/popets-2015-0007>
- Edelman, Benjamin G. and Luca, Michael (2014): "Digital Discrimination: The Case of Airbnb.com (January 10, 2014)". *Harvard Business School NOM Unit Working Paper No. 14-054*. Doi: <http://dx.doi.org/10.2139/ssrn.2377353>
- Epstein, Robert and Robertson, Ronald (2015): "The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections". *PNAS*, 112(33), E4512-E4521. Doi: [doi.org/10.1073/pnas.1419828112](http://doi.org/10.1073/pnas.1419828112)
- Fuller, Matthew (2003): *Behind the blip. Essays on the culture of software*. New York: Autonomedia.
- Hajian, Sara and Domingo-Ferrer, Josep (2013): "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining". *IEEE Transactions on Knowledge and Data Engineering*, 25(7), 1445-1459.
- Koene, Ansgar (2017): "Algorithmic bias: addressing growing concerns". *IEEE Technology and Society Magazine*, 36 (2), 31-32.
- Lévy, Pierre (2013): Le médium algorithmique. Disponible en [http://pierrelevyblog.files.wordpress.com/2013/02/00-le\\_medium\\_algorithmique.pdf](http://pierrelevyblog.files.wordpress.com/2013/02/00-le_medium_algorithmique.pdf)
- Lodemel, Ivar and Trickey, Heather (Eds., 2001): *An offer you can't refuse. Workfare in international perspective*. Bristol: Policy Press.
- McAfee, Andrew and Brynjolfsson, Erik (2012): "Big Data: The Management Revolution". *Harvard Business Review*, October 2012. Disponible en <https://hbr.org/2012/10/big-data-the-management-revolution>

- Mikolov, Tomas; Chen, Kai; Corrado; and Dean, Jeffrey (2013): "Efficient Estimation of Word Representations in Vector Space". Disponible en <https://arxiv.org/abs/1301.3781>
- ONU (2016): Human Rights Council. Thirty-second session. Agenda item 3. Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development.
- Pariser, Eli (2017): *El filtro burbuja*. Madrid/Barcelona: Taurus
- Parisi, Luciana (2013): *Contagious Architecture Computation, Aesthetics and Space*. Cambridge MA (EEUU): MIT Press.
- Rainie, Lee and Anderson, Janna (2017): "Code-Dependent: Pros and Cons of the Algorithm Age". *Pew Research Center, Internet & Technology*, February 2017. Disponible en <http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age>
- Rouvroy, Antoinette et Berns, Thomas (2013): "Gouvernementalité algorithmique et perspectives d'émancipation: le disparate comme condition d'individuation par la relation ? Politique des algorithmes. Les métriques du web". *RESEAUX*, 31(177), 163-196.
- Sweeney, Latanya (2013): "Discrimination in Online Ad Delivery". *ACMqueue - Storage*, 11(3). Doi: 10.1145/2460276.2460278

---

Lucía Benítez Eyzaguirre es formadora y consultora en comunicación, tecnologías e innovación. Periodista, socióloga y realizadora. Profesora de la Universidad de Cádiz. Profesora del Máster de Marketing Digital y del Máster Gestión Estratégica e Innovación en Comunicación. Doctora por la Universidad de Sevilla. Premio Extraordinario de Doctorado de la Universidad de Sevilla y Premio RTVA a la mejor tesis doctoral de Comunicación. Máster en Tecnologías digitales y sociedad del conocimiento. Máster en Inmigración. Experta universitaria en Realización Audiovisual y en Software Libre. Como miembro del grupo de investigación COMPOLÍTICAS investiga las prácticas comunicativas y los usos de la tecnología. En el campo de la Comunicación para el desarrollo participa en proyectos de cooperación de investigación y de enseñanza del periodismo y las nuevas tecnologías en diferentes países de África y América Latina.