

Retos éticos de la ciencia de datos

Antes y



después de una

pandemia

KAREN GONZÁLEZ FERNÁNDEZ, ENRIQUE SIQUEIROS Y LEÓN PALAFOX

«La revolución tecnológica más veloz que ha presenciado la humanidad fue acelerada por la pandemia más grande del siglo XXI. Una vez más tendremos que preguntarnos sobre lo justo y lo injusto hasta ver resuelto lo necesario».

DIMENSIÓN DE LA REVOLUCIÓN INFORMÁTICA

Hablamos de una revolución tecnológica cuando una herramienta cambia drásticamente la organización de los animales humanos: desde el dominio del fuego y el crecimiento cerebral generado por la ingesta de alimentos cocidos, hasta la revolución industrial y nuestro crecimiento demográfico exponencial posibilitado por la medicina moderna. Hoy exploramos las posibilidades de otra transición tecnológica: la revolución informática.

La revolución informática, llamada también 4ta Revolución Industrial, forma parte de un proceso de transformaciones tecnológicas de los últimos 200 años en las que el animal humano sobrepasó sus ritmos naturales, es decir, fue rebasado por su propia herramienta: si la primera y segunda revolución —máquina de vapor, motor de combustión interna y línea de producción— rebasaron la fuerza muscular de animales humanos y domesticados, la tercera y cuarta —automatización, internet, *big data* y *machine learning*— rebasaron su fuerza cognitiva: su capacidad intelectual. ¿Cómo? Mediante la recabación de enormes cantidades de información y el uso de algoritmos computacionales. El caso más patente es la nueva automatización laboral, que va desde el reemplazo de trabajos mecánicos, hasta la toma de decisiones: desde la albañilería y la talacha jurídica hasta las decisiones bursátiles y el otorgamiento de créditos bancarios. El caso más extremo son las disyuntivas mortales de un coche autónomo. ¿Cómo reaccionará mi auto en un accidente de tránsito?

La revolución informática tiene una particularidad: su velocidad. Mientras que las tres «primeras» se dieron en 200 años, y las anteriores en más de un millón, la cuarta empezó hace aproximadamente 30 años, con el inicio de la *www*. No exageran quienes afirman que hoy tu lavadora tiene mayor poder de procesamiento que la nave con la que pisamos la luna. Las llaman tecnologías exponenciales. A este ritmo: ¿quién

puede predecir lo que veremos en cinco años? En la organización social, por ejemplo, percibimos el declive de la autoridad vertical: igual que algunos padres de familia piden ayuda para encender Netflix o el profesor necesita que le conecten el proyector, los gobiernos piden cuentas a los líderes tecnológicos y las monedas comienzan a respaldarse horizontalmente (*blockchain*). Pero a partir de la pandemia del Covid-19, gracias al éxito que mostró el manejo centralizado de datos personales, en controles sanitarios como el de China, antiguas formas de organización vertical y autoridad dictatorial comienzan a persuadir el ojo occidental. Frente a estos giros, la velocidad y la emergencia sanitaria, el individuo, consciente o no, ve amenazada por varios frentes su seguridad, dignidad y libertad.

¿Qué problemáticas humanas acarrea la llamada 4ta Revolución Industrial, acelerada por la volatilidad actual? A continuación, ofreceremos al lector un recorrido de su impacto, tomando como narrativa la estructura básica de una de las últimas disciplinas informáticas, la ciencia de datos o *big data*, encargada de computar mares de información para describir la realidad de manera más precisa y así poder tomar mejores decisiones económicas, políticas y sociales. Dicho rápido, la ciencia de datos consiste en la recabación, almacenamiento, procesamiento y uso de datos, mediante algoritmos de inteligencia artificial, para la toma de decisiones.

RECABACIÓN DE DATOS Y ALCANCES

Desde el pasaje bíblico de Números o el catálogo de las Naves en la *Iliada*, encontramos ejemplos de censos y recabación de datos. Nada nuevo. En las ciencias informáticas, la obtención de datos fue de la mano con el desarrollo de las computadoras. A finales de los 90 cuando surgió el internet y los primeros registros en línea, empezó la necesidad de almacenar datos. Pero si recordamos, nuestra primera computadora

tenía menos capacidad que un USB del Oxxo más cercano. Por ello, cuando las computadoras sencillas se vieron rebasadas, se crearon granjas de servidores y a partir de entonces fue progresando la capacidad de guardar cada vez más información. El presente y futuro de la recabación de datos presenta posibilidades insospechadas: desde la información que diaria y acriticamente compartimos mediante la interacción con nuestros dispositivos móviles, hasta la nueva tecnología de comunicación celular 5G, que promete transmitir datos a velocidades sustancialmente mayores y a una cantidad mucho mayor de dispositivos al mismo tiempo.

Múltiples expertos coinciden en que el futuro del poderío tecnológico recae en la capacidad de desplegar esta red 5G en un país completo, y hoy en día, el único país con las capacidades de hacerlo con la escala necesaria es China, un país con un gobierno fuertemente autoritario y conocido por coartar las libertades civiles.

ALMACENAMIENTO DE LOS DATOS Y PRIVACIDAD

Las nuevas tecnologías en ciencia de datos están generando situaciones paradójicas: una persona puede generar millones de datos con todos los clics que da en una computadora o un celular, y en ese sentido los datos le pertenecen (como el elemento causante de los mismos); sin embargo, esa misma persona no tiene la capacidad de recabar esos datos directamente, ni puede tener un acceso sencillo a los mismos. Además, tampoco cuenta con las herramientas computacionales (algoritmos) para usarlos o interpretarlos. En consecuencia, podemos preguntarnos, ¿esos datos son realmente suyos? ¿Son de quien tiene la tecnología para recabarlos y procesarlos? Responder estas preguntas no es sencillo en el presente contexto.

Las nuevas tecnologías asociadas a la ciencia de datos están obligando a replantear qué se

entiende por privacidad. Desde el nombre hasta los datos biométricos, los identificadores únicos de cada persona se han multiplicado, y muchas instituciones públicas y privadas tienen acceso a estos datos, y se usan para garantizar la seguridad de diversos tipos de operaciones, como las bancarias, por ejemplo. Sin embargo, el reto está en garantizar la privacidad de estos datos a diferentes niveles.

Por un lado, cada institución debe definir para qué puede tener acceso a estos datos y establecer claramente los límites de para qué pueden ser usados y cuánto tiempo podrán tener acceso a los mismos. Una institución financiera puede recurrir a estos datos para garantizar que una persona que quiere realizar una operación bancaria realmente es la dueña de la cuenta en la que se harán los movimientos, por ejemplo; pero los datos que pueden recabar los servicios de salud o las aplicaciones que tratan de prevenir que las personas se enfermen tienen otra función.

Por otro lado, hay que tratar de garantizar que en todas las fases de los procesos vinculados a los datos: generación, recabación y almacenamiento, nadie tendrá acceso a estos para potencialmente hacer un mal uso de los mismos.

Sin embargo, también hay que tener en cuenta que, posiblemente, para que la ciencia de datos pueda desplegar gran parte de su potencial y ser realmente útil para la sociedad, se requiere tener acceso a los datos de una manera que podría ir en

desacuerdo con los lineamientos referentes a la privacidad.

PROCESAMIENTO DE LOS DATOS Y SESGOS INFORMATIVOS

Aunque la recabación de datos sea objetiva en muchos niveles, no deja de estar sesgada, es decir, produce errores sistemáticos o tendencias desproporcionadas a favor o en contra de algo. ¿A qué se debe esto? Hay múltiples razones, algunas de ellas son las siguientes:

no exageran quienes afirman que hoy tu lavadora tiene mayor poder de procesamiento que la nave con la que pisamos la luna.

- Solo tienen poder para generar los datos quienes tienen acceso a la tecnología necesaria para hacerlo: computadoras, teléfonos inteligentes, internet. Así, aunque tengamos millones de datos, son datos referentes únicamente a una parte de la población, y no a toda la población en su conjunto.
- Los algoritmos de redes sociales y de *marketing* tienden a estar diseñados para presentarle a las personas noticias, productos o servicios que estén de acuerdo con sus intereses, valores y necesidades, sin embargo, esto puede generar que las personas se encuentren en los entornos virtuales, en una especie de «cámara de eco», en la que solo refuerzan ciertos puntos de vista, sin ser críticas con ellos o tener acceso a puntos de vista divergentes. Esto se vuelve más peligroso cuando hay grupos con ciertas agendas en las que específicamente se busca generar una tendencia en las opiniones de las personas y que utilizan las herramientas tecnológicas para favorecerla, como sucedió en el caso de Cambridge Analytica, por ejemplo.
- Finalmente, porque describen la realidad: en los procesos de recabación de datos, y la generación y uso de algoritmos, pueden producirse sesgos que responden a diferentes fenómenos, pero es importante notar

```

Training Step: 1999 | totalTraining Step: 1999 |
loss: 0.01591 | time: 0.003sloss: 0.01591 | time: 0
| Adam | epoch: 1000 | loss: | Adam | epoch: 1000 |
0.01591 - acc: 0.9997 -- 0.01591 - acc: 0.9997 --
iter: 16/22 iter: 16/22
Training Step: 2000 | totalTraining Step: 2000 |
loss: 0.01561 | time: 0.006sloss: 0.01561 | time: 0
| Adam | epoch: 1000 | loss: | Adam | epoch: 1000 |
0.01561 - acc: 0.9997 -- 0.01561 - acc: 0.9997 --
iter: 22/22 iter: 22/22

```

```

[[0.004509848542511463[[0.004509848542511463
0.9954901337623596]] 0.9954901337623596]] [[0.004509
0.99549013
[[0.9810173511505127, [[0.9810173511505127, [[0.981017
0.018982617184519768]]0.018982617184519768]]0.01898261

```


frente a la velocidad de la emergencia sanitaria la, el individuo, consciente o no, ve amenazada por varios frentes su seguridad, dignidad y libertad.

que es muy posible que incluso con la ciencia de datos se tienda a replicar los sesgos que ya existían en la sociedad antes de la aparición de las nuevas tecnologías. Se ha estudiado, por ejemplo, que los algoritmos usados en Estados Unidos para evaluar la posibilidad de reducir condenas carcelarias solían presentar sesgos relacionados con cuestiones raciales; lo mismo con algoritmos que usan para evaluar candidatas a ciertos trabajos y que generaban discriminación por género.

Las soluciones a las problemáticas de los sesgos deben proponerse a múltiples niveles. En primer lugar, es muy importante ser conscientes de que la ciencia de datos realmente genera este tipo de problemas y que, por lo mismo, es muy importante estar evaluando constantemente el comportamiento y los resultados que arrojan los algoritmos para poder corregir en donde haga falta.

Sería muy útil que, en la mayor medida posible, se traten de recabar datos que correspondan a todos los diversos sectores de la sociedad y no solo a una parte.

Además, se debería profundizar a nivel educativo, en la importancia del pensamiento crítico que debería ser realizado por todos los ciudadanos, y que ayudaría a comprender la relevancia

de considerar puntos de vista opuestos a los propios, ser capaz de buscar buenas fuentes de información y de reconocer *fake news*, y ser más racional en cuanto a lo que se consume. Por otro lado, desde el ámbito legal debe trabajarse en disposiciones que ayuden a evitar situaciones como las acontecidas en el caso de Cambridge Analytica.

Finalmente, los desarrolladores de algoritmos deben estar monitoreando todo el tiempo los mismos, e introducir modificaciones específicas para evitar sesgos, una vez que estos sean detectados.

USO DE LOS DATOS. CASO DE ESTUDIO, COVID-19

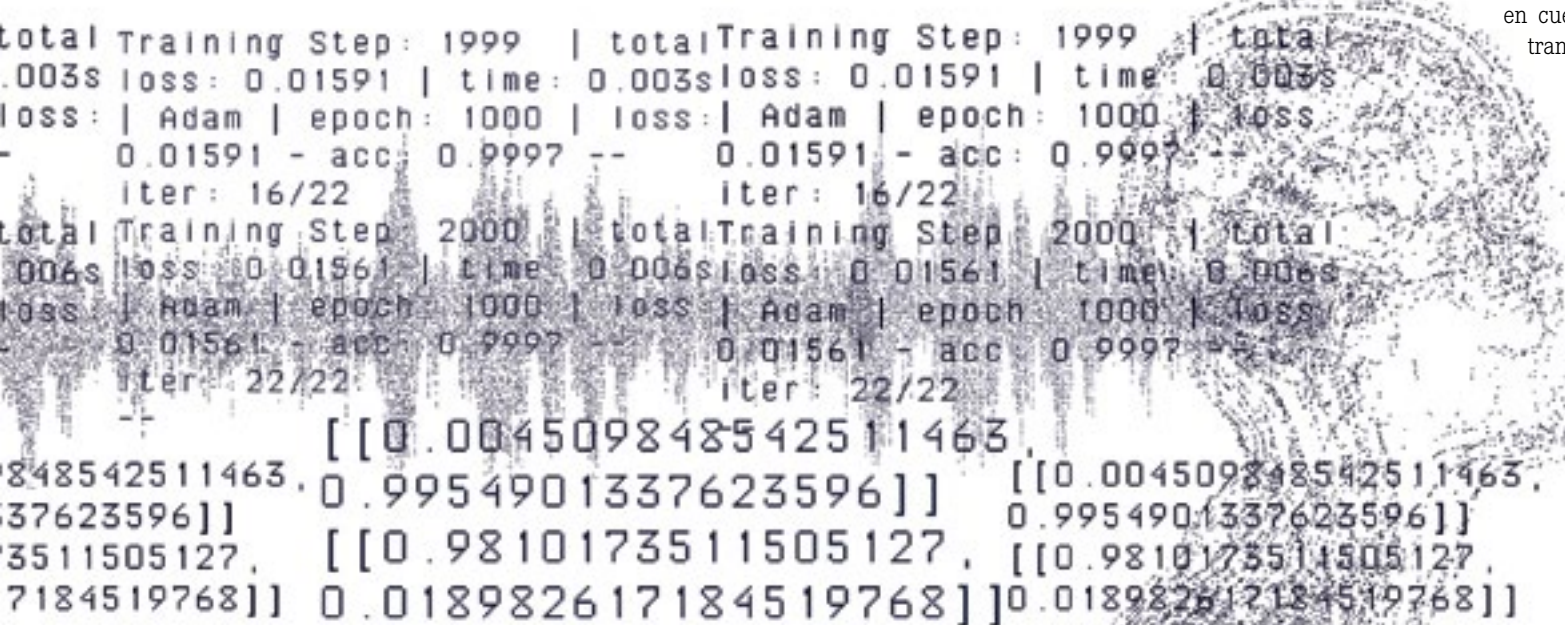
En estos instantes, el mundo está viviendo la pandemia más impactante de nuestra generación. Un virus tipo Corona, que con síntomas similares a los de una influenza, ha reclamado ya más de 25,000 vidas, con más de medio millón de casos a nivel mundial, los cuales día con día se incrementan exponencialmente (datos del 27 de marzo de 2020).

Múltiples países han ordenado cuarentenas obligatorias, se han impuesto toques de queda, hay millones de desempleados y al momento de escritura de este artículo, el Primer Ministro de Inglaterra Boris Johnson se encuentra infectado de la enfermedad.

Estamos viendo, en tiempo real, como países con la experiencia y los recursos para desplegar la ciencia de datos y la inteligencia artificial, la han estado usando para controlar el virus. El contraste entre la adopción de la tecnología y no adoptarla, está convirtiéndose en la diferencia entre la vida y la muerte para cientos de miles de personas.

China, a lo largo de los años, ha instalado una intrincada red de sensores biométricos con el fin de monitorear a su población. Dicha red les permite saber, en tiempo real, dónde se encuentra cada uno de sus habitantes. Sus actividades son medidas, y con base en ellas, se les da un índice

social, dicho índice después es tomado en cuenta para transacciones





el uso de la tecnología, una vez más, está aumentando la brecha de riqueza entre los países desarrollados y los que están en vías de desarrollo.

México podría evitar caer en una recesión aún más profunda, ya que depende fuertemente del comercio informal, así como de las transacciones que se realizan al día con día.

tan complejas como pedir un crédito o tan básicas como comprar un boleto de autobús.

Aumentando su red con sensores térmicos, el gobierno chino es capaz de ver en tiempo real, quién tiene fiebre, y en qué momento se le detonó, ver con quien ha tenido contacto esa persona y con eso, crear uno de los planes de contención más específicos de todo el mundo, donde solo aquellos con síntomas o en contacto directo son puestos en cuarentena, mientras que el resto de la economía puede seguir funcionando de manera normal.

En contraste, tenemos países como México, Argentina, Brasil, donde esta tecnología no es usada, entonces se debe recurrir a medidas más extremas de confinamiento masivo, el cual tiene como resultado la pérdida de empleos, y por ende la caída económica de dichos países. El uso de la tecnología, una vez más, está aumentando la brecha de riqueza entre los países desarrollados y los que están en vías de desarrollo.

CONSENTIMIENTO INFORMADO Y SALUD PÚBLICA

Dentro de todo este esquema de utilizar una red de sensores para salvaguardar a la población, debemos atender a la problemática que mencionamos al inicio. ¿Es válido o ético estar monitoreando constantemente a la población?

La respuesta es difícil y compleja, en un país como México, la población jamás accedería a un tipo de monitoreo o control tan restrictivo como el crédito social de China. Pero si se utilizara una medida así, en el caso de la pandemia,

SESGOS EN LA SALUD

Y la última pregunta que nos atañe es, ¿qué pasa con aquellas personas que por la propia naturaleza de sus trabajos serán aislados? Imaginemos, bajo una vigilancia estricta, las personas cuyo trabajo consiste en moverse alrededor de otros grupos de personas, ellas son las más expuestas, y las que al final serán infectadas, y como tal puede ser que pierdan sus ingresos o sean despedidas

Esto crea, sin lugar a dudas, sesgos al momento de implementar los sistemas de detección de casos, ya que estarán de manera desproporcionada más cargados de personas de bajos ingresos y con poco acceso a la salud y con ello se crea una perversión más dentro de un sistema que lo que busca por encima de todo es la eficiencia.

Frente la revolución informática, muchos confiábamos en esta nueva capacidad de poder definir criterios éticos antes de escribir algoritmos, pensar antes de actuar. Pero la revolución tecnológica más veloz que ha presenciado la humanidad fue acelerada por la pandemia más grande del siglo XXI. Una vez más tendremos que preguntarnos sobre lo justo y lo injusto hasta ver resuelto lo necesario. Solo esperamos que, frente al éxito de gestiones autoritarias y las posibilidades del control tecnológico, se atiendan las urgencias sociales sin poner en riesgo la libertad para pensar por nosotros mismos y en voz alta. </>



Karen González es licenciada en Filosofía por la UP. Maestra y doctora en Filosofía de la Ciencia por la UNAM.

Enrique Siqueiros es licenciado en Filosofía y maestro (c). en Gobierno y Políticas Públicas por la UP.

León Palafox realizó sus estudios de licenciatura en la Facultad de Ingeniería de la UNAM. Es maestro y doctor por la Universidad de Tokio, Japón.

Bibliografía

- Harari, Y. N. (2020, March 20). *Yuval Noah Harari: the World After Coronavirus: Free to read*. Retrieved from <https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75>
- MicroMultiCopter Aero Technology Co., L. (2020). *MMC's drones used in the battle against the new coronavirus outbreak*. Retrieved 28 March 2020 from <https://www.prnewswire.com/news-releases/mmcs-drones-used-in-the-battle-against-the-new-coronavirus-outbreak-301001511.html>
- Mitchell, A. L. (2018). China's Surveillance State Should Scare Everyone. Retrieved 28 March 2020, from <https://www.theatlantic.com/international/archive/2018/02/china-surveillance/552203/>
- Oppenheimer, A. (2018). *Sálvese quien pueda. El futuro del trabajo y los trabajos del futuro*. Debate.
- Rees, C. (2014). Who Owns our Data? *Computer Law and Security Review*, 30, pp.75-79.
- Richterich, A. (2018). Big Data: Ethical Debates. *The Big Data Agenda. Data Ethics and Critical Data Studies*. (pp. 33-51). University of Westminster Press. <https://doi.org/10.16997/book14.c>
- Wrangham R. (2009). *Catching Fire: How Cooking Made us Human*. Perseus Books Group.

istmo[®]

liderazgo con valores

Forma parte
de esta **comunidad**

¡Suscríbete!

\$480
anual



Código de promoción **IST20**



CONTACTO
mdominguez@ipade.mx
55 1250 1616

Pregunta por nuestras promociones