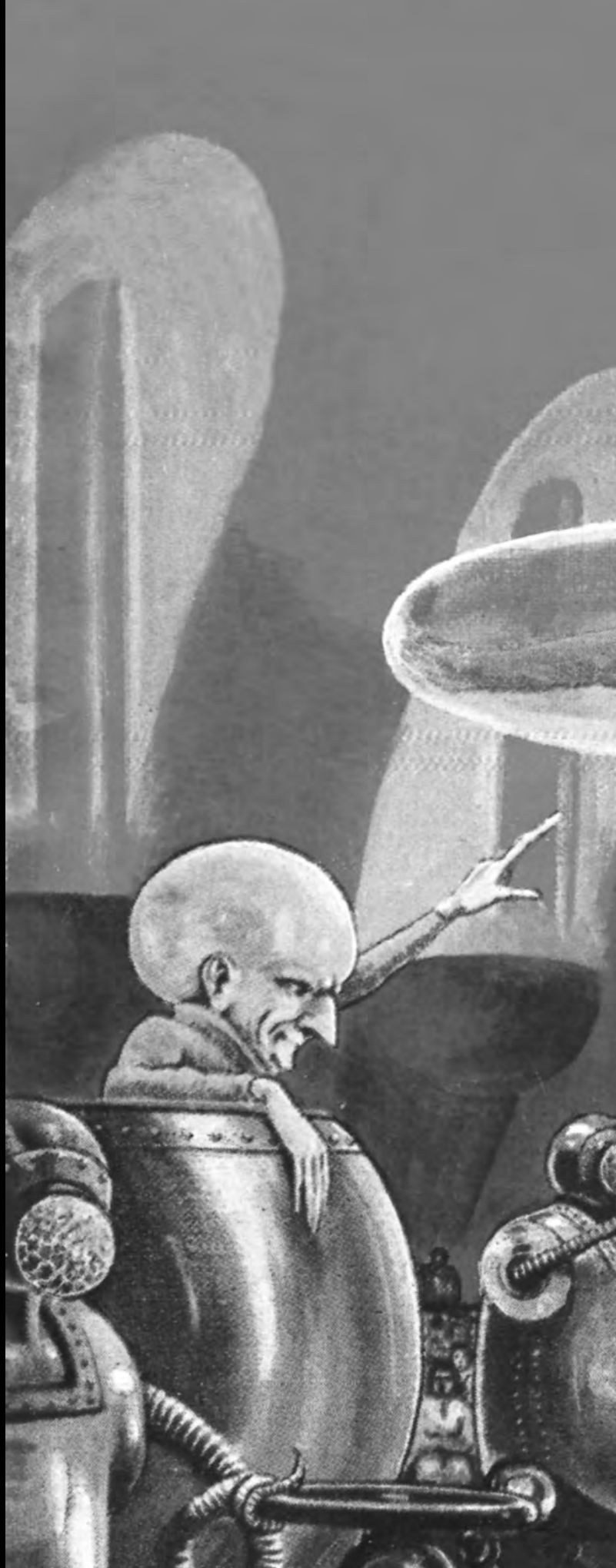


Cuenta la leyenda que cuando el ejército romano tomó por asalto la ciudad de Siracusa, en el siglo III antes de nuestra era, el matemático Arquímedes estaba inclinado sobre la arena dibujando figuras, seguramente embebido en algún problema abstracto; era tal su concentración que no se percató del soldado romano que le ordenó ponerse de pie. Se cuenta que las últimas palabras del sabio griego fueron $\mu\omicron\nu\ \tau\omicron\upsilon\varsigma\ \chi\upsilon\lambda\lambda\omicron\upsilon\varsigma!$, algo así como “¡no me muevas los círculos!”. La figura del científico que, absorto en su trabajo intelectual, falla en reconocer las circunstancias a su alrededor —el cambio de orden político, la vulnerabilidad de su ciudad, las acciones de la gente que lo rodea— podría servir en este momento de la historia como una elocuente parábola para la comunidad científica, particularmente aquellos que participan en el desarrollo de la llamada inteligencia artificial.

Dicha tecnología ya está entre nosotros: nuestros teléfonos celulares, el manejo de nuestros datos por parte del gobierno y las corporaciones industriales. Ya impacta nuestras vidas pues es una herramienta de control social, represión, de inducción al voto y de incitación al consumo de bienes inútiles. En este momento en el que está muy en boga su estudio y desarrollo, se vuelve imperativo reflexionar acerca de las implicaciones sociales que tiene el uso de nuestros algoritmos, los puntos ciegos que podrían tener y las reconfiguraciones de poder que resultarán de su adopción.

Y lo hizo a su imagen y semejanza...

La totalidad de los pueblos del mundo, o casi, tienen mitos acerca de la creación, y aunque varían en sus detalles específicos, todos tienen en común que un dios o varios forman el primer humano a partir de materia inerte y le insuflan la vida mediante un acto de poder divino. Así tenemos que los dioses originarios de los mayas crean a los humanos después de varios ensayos fallidos; primero usando como material el barro, pero el resultado es una humanidad débil y con cuerpos quebradizos; después intentan con madera, pero el resultado tampoco es bueno y, por último, tienen éxito cuando fabrican gente de maíz. En el viejo testamento se afirma que Jehová creó el primer hombre, Adán, empleando polvo como materia prima y le da una compañera, Eva, usando una de sus costillas. En la mitología nórdica, Odín, Lódurr y Höfnir paseaban por la playa y encontraron dos troncos con formas humanoides, Odín les insufla el soplo vital mientras que sus compañeros los dotan de capacidades mentales, entre ellas la razón, que los hace humanos, y de esa manera se origina la primera pareja: Ask y Embla.



Marisol Flores Garrido

La muerte de Arquímedes

implicaciones sociales de
la inteligencia artificial y
el manejo de datos masivos





Todos estos mitos tienen algo en común: la deidad construye los humanos usando una base material inerte —madera, barro, maíz— y después los anima mediante su poder o soplo divino, otorgándoles la vida.

La humanidad ha querido imitar a los dioses y, desde tiempos remotos, imaginaba autómatas que se comportaban como humanos. Está Talos de Creta que, según la mitología griega, era un humanoide gigantesco de bronce que protegía a Europa, la madre del rey Minos, de invasores y forajidos; la misma mitología menciona que Hefesto, el dios de los herreros y los metales, tenía muchos autómatas a su servicio en el trabajo de su forja. El mundo oriental no se quedaba atrás; en la China del siglo tercero de nuestra época, Lie Zi habla en su libro de Yan Shi, un ingeniero que construyó un hombre mecánico que hablaba y se movía de tal manera que se le confundía con cualquier persona.

En los tiempos del racionalismo cartesiano resurge el empeño de crear autómatas, y se fabricaron ingeniosos dispositivos de forma humanoide capaces de jugar ajedrez o tocar algún instrumento musical. En aquel entonces se hablaba del alma, del soplo divino o de la esencia incorpórea. El mismo Descartes aceptaba la existencia del alma pero no le atribuía orígenes divinos, lo que le causó problemas con la Iglesia. La reacción romántica del siglo XIX se aparta de la

razón y su instrumento que es la ciencia, demonizando los intentos humanos de fabricar autómatas; nadie mejor para ilustrar esta tendencia que Mary Shelley con su famosísima obra *Frankenstein o el moderno Prometeo*.

La inteligencia artificial

En 1956, John McCarthy convoca a una pléyade de científicos a un taller en el prestigiado Dartmouth College, un centro de educación superior perteneciente a la famosa Ivy League estadounidense. A dicha reunión asistieron los científicos más notables de la Unión Americana en el naciente campo de la cibernética, entre ellos, el ahora tristemente célebre Marvin Minsky.

El propósito de la reunión era fijar líneas de investigación para la naciente ciencia de la computación. El uso de computadoras programables comenzaba a generalizarse y los asistentes a la reunión se dieron a la tarea de fijar lo que serían las líneas de investigación futura: aprendizaje automático, reconocimiento sensorial y del habla, identificación de patrones, y muchas otras tareas que se englobaron bajo la denominación de inteligencia artificial.

A partir de ese momento se popularizó dicho término y, con muchos altibajos, durante la segunda mitad de los

ochentas, su avance, así como el de su soporte metodológico, el cómputo neuronal, ha sido imparable. Hoy día nuestra vida está rodeada de dispositivos que hacen uso intensivo de ella sin que seamos plenamente conscientes.

Aprendizaje automático

El aprendizaje automático es una rama de la inteligencia artificial que se enfoca en el diseño de algoritmos que realizan tareas específicas, como clasificar y agrupar, utilizando solamente datos como entrada. Aunque existen diversos tipos de aprendizaje automático, para nuestro tema nos concentraremos en el llamado “aprendizaje supervisado”, orientado a tareas de predicción o reconocimiento de patrones

Para plantear una tarea de aprendizaje supervisado necesitamos dos cosas fundamentales: 1) tener claridad en cuanto a las clases en las cuales queremos ubicar los objetos a clasificar; y 2) contar con un número grande de objetos cuya clase se conoce previamente y que están, cada uno de ellos, caracterizados por un vector o una lista de rasgos.

Supongamos que nuestra tarea consiste en determinar si una persona tiene cierta enfermedad o no; en este caso, las clases o categorías en las que se va a clasificar serán simplemente “estar enfermo” y “no estar enfermo”, y enseguida necesitaremos los datos de muchas personas cuya pertenencia a una de las dos clases se conozca de antemano, información que provee un especialista, en este ejemplo, un médico. Cada persona debe estar representada por ciertos atributos o rasgos que consideremos relevantes para la predicción que queremos efectuar; en este caso podrían ser la edad, la temperatura del cuerpo, el ritmo cardíaco, la presión arterial y otras mediciones relevantes.

De esta manera, cada sujeto tendrá un vector de rasgos asociado y la lista de las personas con todo y atributos formarán una matriz que llamaremos “base de conocimiento”. También supondremos que ordenamos los renglones de la matriz de manera que en la parte superior que-

den los individuos enfermos y en la de abajo los no enfermos. El proceso de asignarle una de las dos categorías a un recién arribado es justamente a lo que llamamos “clasificación” o “reconocimiento de patrones”.

Aunque existe una variedad de enfoques mediante los cuales se puede efectuar dicho proceso —redes de neuronas artificiales, enfoques bayesianos, métodos de regresión múltiple— ninguno, por muy sofisticado que sea, es suficiente por sí solo: la calidad del resultado depende fuertemente de los datos con los que se cuente. Esto explica el valor que tiene una buena base de conocimiento para quienes buscan desarrollar algoritmos de aprendizaje automático, las innumerables estrategias de todo tipo que las corporaciones emplean para capturar datos de sus usuarios y el surgimiento de expresiones (muy cuestionables) como “los datos son el nuevo petróleo”.

Esto ayuda también a desmitificar algoritmos que reciben mucha atención por parte de los medios de comunicación; por ejemplo, la próxima vez que escuchemos de un algoritmo que detecta “sarcasmo” en *twitter*, debemos pensar que alguien encontró una manera de representar mu-

chísimos *tweets* mediante un conjunto de atributos y encontró también la manera para relacionar los *tweets* con la etiqueta “sarcasmo” o “no sarcasmo” previamente establecida. Decir que las máquinas comprendan el sarcasmo es un abuso monumental propio de los medios sensacionalistas.

Por otro lado, los resultados a los que se puede llegar mediante la combinación de un volumen grande de datos y herramientas matemáticas adecuadas son verdaderamente impresionantes. Hemos visto desde logros en el área de la salud y traducciones de texto sobresalientes, hasta videos absolutamente realistas de un falso Barack Obama. De hecho, los resultados son tan buenos que han llevado a una parte de la comunidad dedicada a este campo a una confianza exacerbada, como la expresada en un artículo publicado en la revista *Wired* hace algunos años: “éste es un mundo en donde una cantidad masiva de datos y matemáticas aplicadas reemplazan cualquier otra herramienta que



podiera haberse utilizado. Olvide cualquier teoría del comportamiento humano, desde lingüística hasta sociología. Olvide la taxonomía, la ontología y la psicología. ¿Quién sabe por qué la gente hace lo que hace? El punto es que lo hacen y nosotros podemos registrarlo y medirlo con una fidelidad sin precedentes. Con suficientes datos, los números hablan por sí solos". Y es en esta confianza desmedida en donde radica uno de los grandes riesgos de la inteligencia artificial.

Peligros de la inteligencia artificial

Los cuestionamientos al desarrollo de esta tecnología han estado presentes desde su inicio. Turing, en el artículo de 1950 que llegaría a ser un clásico, incluye ya una discusión sobre objeciones que la gente podría tener ante la idea de máquinas inteligentes. Sea porque los humanos establecimos que nuestra habilidad para razonar es el rasgo que nos hace únicos y diferentes a las demás especies o porque la idea de máquinas que piensan conjura inmediatamente en nuestra imaginación futuros distópicos vistos en películas, los avances en dicho campo despiertan preguntas sobre el futuro: ¿cómo será ese espacio compartido con máquinas en todo tipo de tareas?

Para bien o para mal nuestra generación atestiguará una batalla entre Estados Unidos y China por el liderazgo en el terreno de la inteligencia artificial, cada semana vemos algún avance con el que tales potencias buscan posicionarse como líder único: China anuncia que implementará un sistema para calificar a sus ciudadanos y rastrear su confiabilidad, Estados Unidos avisa que a partir de 2025 el *check-in* en todos sus aeropuertos será empleando el reconocimiento facial, luego China hace público que no sólo sus aeropuertos sino que también sus agentes policíacos tendrán lentes especiales con herramientas incorporadas para el reconocimiento facial y así sucesivamente. Esta carrera es preocupante porque no deja tiempo para, ya no digamos regular, sino llevar a cabo un ejercicio de prospectiva que busque qué podría salir mal, qué puntos ciegos podría tener toda esta tecnología y cuales serían sus consecuencias negativas. El neoliberalismo no tiene tiempo para la reflexión.

Aun peor, muchas personas no parecen entender la gravedad de estar atrapados en una batalla que requiere nuestros datos para su desarrollo ni las consecuencias que puede tener incorporar esta tecnología a la vida diaria sin estar organizados de una manera que ponga en la balanza la creciente asimetría entre quienes producen los datos y quie-





nes los usan. Aunque en el imaginario colectivo la frase “riesgos de la inteligencia artificial” evoca batallas imaginarias contra robots que adquirieron conciencia, hay aspectos de ésta mucho menos taquilleros y más peligrosos en el corto plazo.

Afortunadamente, la importancia de cuestionar los usos y modelos de dicha tecnología está recibiendo cada vez más atención y se ha visto traducida en la creación de distintos grupos orientados a explorar su impacto social; entre ellos se destaca *AI Now*, enfocado a la discusión ética en torno a la creación y el uso de sus innovaciones, que ha propuesto el análisis de tres áreas importantes: derechos y libertades, trabajo y automatización, y sesgo algorítmico.

Derechos y libertades

Los algoritmos empleados en este campo utilizan como insumo grandes cantidades de datos, lo cual incrementa el valor de estos últimos y motiva a que cualquier interesado en incorporar en su área el uso de cómputo inteligente busque mecanismos para recolectar información y conformar bases de datos. Además, algunos de estos algoritmos, a su vez, generan datos; aplicaciones como *Waze*, por ejemplo, necesitan datos para funcionar pero también recolectan información de los usuarios para incrementar su base y mejorar su funcionamiento.

El aumento en el volumen de datos generados, la importancia que tienen y la forma en que pueden capitalizarse propician prácticas poco éticas por parte de las empresas y vuelve urgente la discusión sobre temas como la privacidad y la asimetría entre quienes generan los datos y quienes se benefician de su uso. Consideremos, por ejemplo, escándalos que periódicamente llegan a los medios de comunicación: en 2016 *Facebook* recolectó durante meses lo que la gente escribía en la plataforma y que, en el último momento, decidía no publicar; la compañía pretendía generar un conjunto de datos que permitiera estudiar la autocensura. En el mismo año, algunos periodistas del grupo ProPublica denunciaron que la compañía tiene un sistema de categorización de usuarios con etiquetas muy específicas, algunas con sólo dos o tres personas en la categoría, y no tiene reparos en ofrecer a clientes potenciales la identificación de perfiles con características como: “odia a judíos”. Más allá de la ausencia de lineamientos éticos en la compañía, el nivel de especificidad y vigilancia que queda de manifiesto es tan preocupante como la falta de regulación para grandes corporaciones.



¿A quién deberían pertenecer los datos? ¿De qué manera nos vemos afectados por todas estas actividades que lucran con nuestra experiencia de vida? ¿Debería permitirse que un empleador potencial tenga acceso a los datos de un solicitante para ver si su perfil es pertinente con el empleo que ofrece? ¿Qué pasa si toda esa información del perfil permite inferir creencias y manipular opiniones, por ejemplo, respecto del voto? ¿Qué pasa con los mecanismos basados en marcadores biométricos que se está tratando de incorporar en aeropuertos y vías públicas? Queda claro que hay riesgos asociados al uso de algoritmos que deberían analizarse a profundidad y que será necesario encontrar la forma de regular desde las leyes el uso de algunas de tales herramientas.

Trabajo y automatización

A medida que crece la lista de tareas que es posible automatizar, inevitablemente vienen las preguntas sobre el nivel de reorganización laboral que la inteligencia artificial tendrá en consecuencia. La disminución esperada en la demanda de trabajo humano es un motivo de preocupación que no es menor: aunque en la retórica de muchas empresas los humanos no serán reemplazados sino “liberados” de tareas repetitivas, hay reportes que sugieren un fuerte interés por parte de la industria privada en recortar, tanto y tan rápido como sea posible, el personal con el que cuentan. Ganar la carrera no oficial que existe en el mercado para alcanzar mayor eficiencia (mayor ganancia) con menos empleados (menores costos) mediante la automatización es un logro que viene acompañado de una importante ganancia económica.

El debate sobre la incorporación de máquinas al entorno laboral no es nuevo y hay autores que vislumbran incluso el surgimiento de un neoludismo. Más allá de la discu-

sión, queda claro que el incremento de tareas que pueden automatizarse traerá consigo un reacomodo en el entorno laboral y será necesario tomar acciones para mitigar el efecto negativo que la transición hacia esta industria 4.0 podría tener entre los sectores más vulnerables. Más aún, la discusión debería extenderse para abarcar la forma en que la incorporación de algoritmos para automatizar decisiones y procesos impactará directamente la relación entre el empleador y el empleado, las dinámicas laborales y de poder al interior de las compañías y la responsabilidad profesional.

Sesgo algorítmico

Los algoritmos, con la capacidad de utilizar un volumen de información inimaginable para un humano, reconocer patrones desapercibidos para nosotros y orientar una decisión con absoluta fidelidad a criterios preestablecidos, se presentan como una alternativa que parece adecuada para lidiar con los errores sistemáticos que exhibimos los humanos en la toma de decisiones. Un algoritmo que decide sobre una contratación o una sentencia judicial no será víctima del efecto halo, ni tomará en cuenta el género, la vestimenta o el color de la piel de los involucrados.

Automatizar algunas decisiones se esboza como la posibilidad de contribuir a una sociedad más justa y equitativa; sin embargo, de cuando en cuando nos topamos con noticias que nos hacen dudar: el *bot* que Microsoft lanzó en Twitter en 2016 y tuvo que ser desactivado porque en sólo unas horas aprendió a hacer comentarios racistas o el sistema de reconocimiento de imágenes de Google que etiquetaba per-



sonas negras como “gorilas” y respondía a la petición de búsqueda “cabello profesional” con imágenes de mujeres blancas de cabello lacio.

¿Qué puede decirse de estos casos de “malfuncionamiento” detectado en algoritmos? ¿Por qué suceden? ¿Son casos aislados? ¿Qué nos dicen sobre la confianza que deberíamos depositar en el uso de inteligencia artificial en temas importantes?

La expresión “sesgo algorítmico” (en inglés *machine bias*) ha cobrado fuerza en los últimos años para señalar en los algoritmos un comportamiento no deseado —aunque los problemas que han agrupado bajo dicho término son de naturaleza distinta. El Instituto AI Now, por ejemplo, utiliza tal expresión para señalar casos en los que un algoritmo podría perjudicar a un sector específico de la sociedad, mientras el grupo ProPublica lo ha empleado para abordar decisiones poco éticas implementadas en algoritmos por ciertas empresas para obtener una ganancia comercial; por ejemplo, en septiembre de 2016, ProPublica reportó la manipulación por parte de Amazon de las opciones de artículos en venta que muestra a sus usuarios con la intención de favorecer aquellas que representan mayor ganancia para la empresa.

Los siguientes ejemplos ilustran otras situaciones que han sido reportadas bajo el término sesgo algorítmico y ayudan a comprender la importancia que el tema ha cobrado.

Datos excluyentes. Puesto que los datos son la base sobre la cual se desarrollan tareas de aprendizaje automático, un algoritmo está condenado a reproducir lo que encuentra en ellos. Zach Weinersmith, autor del comic *Saturday Morning Breakfast Cereal*, representa muy bien esta situación en una de sus ilustraciones: un algoritmo que utilizara datos sobre las batallas que se han librado en nuestro planeta, intentaría reproducir la relación detectada entre ganar una batalla y utilizar lanzas y piedras; el ejemplo puede parecer simple pero numerosos casos reportados como sesgo algorítmico tienen su origen en un error del tipo mostrado en el comic.

Tomemos el caso del sistema de recomendación de YouTube, que busca maximizar el tiempo que un usuario pasa viendo contenido en la plataforma y, puesto que detecta que muchos usuarios que ven videos relacionados con teorías de conspiración pasan horas en la plataforma, el sistema incorpora este tipo de videos entre sus principales recomendaciones.

Un algoritmo refleja los datos que lo alimentan. Datos sesgados se traducen en resultados sesgados, como se ha documentado extensamente en el área de visión computacional. Así lo muestra el caso reportado por Joy Buolamwi-

ni, una empleada del MIT Media Lab y fundadora de la Algorithmic Justice League; afroamericana, Joy se topó con algunos algoritmos de reconocimiento facial que presentaban deficiencias para identificar su rostro y, al investigar la causa del problema, encontró que muchos de éstos se construyen utilizando bases de datos que fallan en representar apropiadamente a la población, lo cual ilustra claramente la base de datos *Labeled Faces in the Wild*, que consta de 30 281 imágenes de rostros recolectadas en medios de co-



municación y cualquiera puede consultar en la red, y es ampliamente utilizada para entrenar algoritmos de aprendizaje automático relacionados con reconocimiento facial, cuyo análisis reveló que 77.5% de las caras que la constituyen son de hombres y 83.5% de personas blancas. A la luz de esto, resulta poco sorprendente que un algoritmo entrenado con tales datos presente errores cuando se usa para el rostro de una mujer negra; de hecho, se ha reportado incluso que para personas con tez oscura la probabilidad de ser atropelladas por un vehículo autónomo es mayor.

En este tipo de casos generalmente encontramos problemas metodológicos que transforman las deficiencias de los datos de entrenamiento en algoritmos que tienen un mal funcionamiento para sectores específicos de la población; muchos pueden evitarse igualmente mejorando el proceso de construcción de la base de datos y poniendo particular atención en que los datos que se usan representen adecuadamente a la población. Desafortunadamente, no todos los casos de sesgo algorítmico se arreglan balanceando los datos.

Las mujeres en la tecnología. Aun cuando las personas a cargo de desarrollar un algoritmo sean cuidadosas en utilizar datos que incluyan distintos sectores de la población, es posible que los datos mismos capturen procesos históricos marcados por la desigualdad.

Meredith Whittaker ilustra este hecho refiriendo el caso de la tarjetas *Shirley* utilizadas durante décadas por Kodak para calibrar el tono de piel en las fotografías. Estas imágenes, nombradas así en honor a la primer modelo en aparecer en ellas, muestran a una mujer caucásica con algún vestido colorido y funcionan como referencia para el balance de color. Dado este proceso de calibración, las imágenes

capturadas por medio de esta tecnología, hasta mediados de los noventas, muestran una definición diferente para pieles blancas y oscuras. Por tanto, si entrenáramos un algoritmo utilizando fotografías capturadas durante este periodo de la historia, tendríamos una diferencia en la representación que no se resuelve balanceando los datos y que refleja relaciones de poder dominantes en la sociedad en donde se crearon los datos.

En una línea semejante, con datos que reflejan desigualdad, tenemos el caso del algoritmo de LinkedIn, la red social orientada a conectar profesionistas y empleadores potenciales, cuyo algoritmo de búsqueda priorizaba nombres masculinos y sugería incluso “corregir” búsquedas relacionadas con mujeres (“tal vez quisiste decir Stephen en vez de Stephanie”); o el algoritmo utilizado por el sistema de justicia estadounidense para estimar la probabilidad de que una persona cometa un crimen en el futuro, que etiquetaba a personas negras con un riesgo mucho más elevado respecto de personas blancas.

Ambos casos reproducen el error mostrado en el comic mencionado: el algoritmo hace predicciones a partir de datos históricos que capturan condiciones de desigualdad y reproduce lo encontrado. Aun más, el comportamiento de los algoritmos contribuye a perpetuar la desigualdad, por ejemplo, favoreciendo la promoción de un hombre en vez de una mujer para determinado empleo y creando un círculo vicioso, pues ese empleo se convertirá en un dato para alimentar futuras decisiones del algoritmo.

Google, niños y niñas. Finalmente, consideremos un caso que circuló en redes sociales en marzo de 2019: capturas de pantalla hechas desde el teléfono celular de una persona en la ciudad de México que muestran las sugerencias de búsqueda hechas por Google a partir de la palabra “niños” y “niñas”; el contraste es escandaloso, pues en el caso de “niñas” las principales sugerencias de búsqueda tienen un matiz sexual que no aparece en el otro caso.

La primera reacción de quienes observan este ejemplo es suponer que el comportamiento del algoritmo no hace más que reflejar el volumen de búsqueda asociado con cada sugerencia y, en consecuencia, nosotros, los usuarios, somos los responsables absolutos del contraste observado. Sin embargo, aunque definitivamente tenemos problemas importantes de género en nuestra sociedad, quedarnos con esta primera idea es pasar por alto un agente importante: el algoritmo mismo de Google, el cual posibilita tales diferencias en los resultados y las permite en la medida que le re-





presentan un beneficio comercial. La empresa tuvo una respuesta inmediata en el momento que el tema cobró fuerza en las redes sociales: bloqueó rápidamente las sugerencias “inapropiadas” e, incluso, agregó un botón para reportar sugerencias de búsquedas ofensivas; cualquiera diría que ya había una ruta de acción preparada para una situación que, además, la compañía había enfrentado (y resuelto) varios años atrás en Estados Unidos.

Safiya Noble presenta una importante discusión sobre el tema en su libro *Algorithms of Oppression*. Si no es una cuestión técnica, ¿a qué puede obedecer la ventana de tiempo entre tener la solución a un problema detectado e implementarla? Aunque mucha gente piensa en el buscador de Google como en un mayordomo amable que nos permite encontrar respuestas, estamos ante una compañía privada cuyo negocio se basa en la venta de publicidad. Hace casi veinte años se hablaba con admiración en los círculos académicos del algoritmo de PageRank y se especulaba sobre su funcionamiento basado puramente en la estructura de red de internet. Muchas cosas han sucedido al interior de la compañía desde entonces y en este momento el ordenamiento que muestra como resultado a una petición de búsqueda obedece a decisiones que no son abiertas al público y que, especulamos, se guían por una combinación de sugerencias úti-

les para el usuario y de sugerencias que representan una ganancia para la empresa.

Con base en los casos mostrados podríamos proponer que la expresión “sesgo algorítmico” hace referencia a una desviación sistemática, no anticipada por el usuario final, en los resultados de un algoritmo, la cual puede deberse a problemas metodológicos relacionados con los datos o la representación de un grupo social que se encuentra asociado a procesos históricos, o bien fue inducida por el grupo a cargo del desarrollo del algoritmo sin que esto sea claro para el usuario.

A partir de los numerosos casos reportados de este tipo de problemas, los sectores involucrados en el desarrollo de inteligencia artificial —academia, industria y gobierno— se han enfocado en crear líneas de análisis alrededor de los algoritmos y su impacto en la sociedad bajo nombres cada vez más diversos y específicos —desde “justicia algorítmica” hasta “equidad en los datos”.

Nuestras herramientas en tiempos oscuros

Que el análisis crítico de los algoritmos cobre importancia y se vuelva un tema central al interior de la comunidad es una buena noticia. Se requiere miradas diversas para iden-



tificar posibles puntos ciegos en nuestra tecnología, así como herramientas que van más allá de las matemáticas o las ciencias computacionales para desarrollar una inteligencia artificial benéfica. Independientemente de los casos de sesgo algorítmico, en los últimos años hemos visto publicaciones científicas que exhiben un grado preocupante de confusión respecto de los problemas que pueden abordarse mediante esta tecnología.

Apenas en noviembre de 2016, un grupo de investigación de la Universidad de Shangái Jiao Tong publicó un artículo en el que afirma ser capaz de usar una red neuronal convolucional para determinar, a partir de una fotografía del rostro, si una persona es un criminal, el cual resulta alarmante no sólo por su potencial para ser mal usado, sino por la ignorancia que exhiben los autores respecto de un problema que dista mucho de ser nuevo y a la candidez con que proponen abordarlo.

Determinar la criminalidad de una persona a partir de rasgos en su fisonomía es exactamente la idea propuesta por Cesare Lombroso en el siglo XIX, la cual deriva de la “frenología” y ha sido ampliamente discutida y, al final, desechada; en su momento sirvió para revestir de ciencia y objetividad algunas creencias de la época que justificaban la supremacía

de un grupo de poder sobre el resto de las personas. Cuando las ideas de Lombroso llegaron a México, por ejemplo, científicos nacionales de la época utilizaron dicha teoría para establecer que los indígenas en nuestro país tenían un rostro que denotaba miseria, abandono, falta de inteligencia y violencia innata, entre otras cosas.

Es importante destacar que en el caso de esta publicación de 2016 no tenemos un problema propiamente de sesgo algorítmico o de un dilema ético sino, mucho más grave, una alarmante confusión epistémica que lleva a los autores a revivir una pseudociencia bajo el cobijo de herramientas sofisticadas de inteligencia artificial.

Por supuesto, en una sociedad que es capaz de revivir la idea de que la Tierra es plana hay terreno fértil suficiente para sembrar la idea de que esta tecnología, dotada de un procedimiento muy sofisticado y científico, puede usarse para determinar la personalidad (¡o la criminalidad!) de las personas a partir de su rostro. Y en una sociedad donde impera un sistema capitalista voraz no falta quien aproveche la situación: páginas como Faception intentan vender frenología moderna explicando la utilidad que podría tener en la investigación de crímenes o en puntos de control en las fronteras. Para estar a tono con la conversación en la comunidad dedicada a este campo, la empresa tiene incluso un apartado donde discute ética y justicia en sus algoritmos.

Si en algún momento ha sido apremiante un diálogo entre las ciencias exactas y las humanidades, es ahora.

Por un desarrollo tecnológico responsable


Con tantos intereses económicos en juego, es casi imposible que se dé una desaceleración en el desarrollo de la inteligencia artificial, aun si son muchas las voces que aconsejan detener la implementación de ciertas herramientas hasta contar con un análisis a profundidad sobre sus posibles consecuencias, como es el caso del reconocimiento facial en lugares públicos.

¿Qué hacer? Existen voces que sugieren establecer un juramento hipocrático para matemáticos, computólogos y especialistas involucrados en el desarrollo de tecnología, como un llamado a reflexionar en las posibles aplicaciones que su trabajo podría tener y a centrarse en el bienestar de la sociedad —porque, seamos claros en esto, la inteligencia artificial y el manejo de un volumen masivo de datos puede traer grandes beneficios a la sociedad, lo ha hecho ya en áreas como la salud pública. Paradójicamente, junto con las amenazas que puede tener su mal uso, ésta provee también

herramientas esperanzadoras para enfrentar problemas inminentes en nuestro planeta.

No se trata de frenar el desarrollo de esta tecnología, sino de mantener una postura crítica frente a las herramientas que posibilita. Será importante que quienes participan en su desarrollo mantengan las cosas en perspectiva, integrando en su ejercicio investigativo un diálogo con especialistas de otras disciplinas; establecer grupos de trabajo interdisciplinarios será fundamental para explorar la dimensión social, cultural y política de las herramientas que se están generando y proponer estrategias enfocadas en mitigar los

efectos que la adopción de tales tecnologías podría tener en las minorías y los grupos más vulnerables.

Asimismo, es necesario recordar que nuestros modelos comienzan desde los datos, pues decidir esquemas de representación ya establece una relación de poder. Por sofisticados que sean, nuestros modelos siempre reflejan nuestro punto de vista y, en consecuencia, capturan nuestra visión del mundo. Tener esto en cuenta nos recuerda la importancia de mantener en todo momento un vigilancia epistémica y un punto de vista crítico que reconozca las limitaciones propias. 



Marisol Flores Garrido

Escuela Nacional de Estudios Superiores
Unidad Morelia,
Universidad Nacional Autónoma de México.

REFERENCIAS BIBLIOGRÁFICAS

Big Data Increases Inequality and Threatens Democracy. Broadway Books, Nueva York.

Broussard, Meredith. 2019. *Artificial Unintelligence: How Computers Misunderstand the World*. MIT Press, Cambridge Mass.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, Nueva York.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Algorithms Inequality*. Crown Publishing Group, Nueva York.

Kearns, Michael y Aaron Roth. 2019. *The Ethical Algorithm: The Science of Socially Aware Algorithms*. Oxford University Press, Oxford.

IMÁGENES

P. 6: portada de *Astounding Science Fiction*, vol. 47, núm. 6, 1951; portada de *Eye in the Sky*, Philip K. Dick, Ace Books, 1957. P. 7: Bob Pepper, portada de *More Than Human*, Ballantine Books, 1970. P. 8: portada de *Amazing Stories: Slaves of the Crystal Brain*, vol. 24, núm. 5, 1950. P. 10: Malcolm Smith, *Imagination Stories of Science and Fantasy*, 1951; Virgil Finlay, *Famous Fantastic Mysteries*, 1942. P. 11: portada de *Astounding Science Fiction*, vol. 47, núm. 4, 1951. P. 12: portada de *Startling Stories*, vol. 3, núm. 2, 1940. P. 13: portada de *Marvel Science Stories: Avengers from Space*, vol. 1, núm. 1, 1938; portada de *Captain Future, Man of Tomorrow: The Days of Creation*, 1944. P. 14: Dave Meltzer, portada del libro *In the Enclosure*, Barry N. Malzberg, Avon Science Fiction, 1973.

AGRADECIMIENTOS

Agradece el apoyo del proyecto PAPIIT IA107318.

SOCIAL IMPLICATIONS OF ARTIFICIAL INTELLIGENCE AND BIG DATA MANAGEMENT

Palabras clave. Inteligencia artificial, ética científica, sociedad, tecnología, aprendizaje de máquina, cómputo neuronal.

Key words. Artificial intelligence, scientific ethics, society, technology, machine learning, neural computation.

Resumen. A medida que la Inteligencia Artificial gana terreno en diferentes áreas del conocimiento, permea actividades cotidianas y adquiere el potencial de impactar vidas, se vuelve importante reflexionar en las implicaciones sociales que tiene el uso de nuestros algoritmos y en los puntos ciegos que estos podrían tener. En este artículo buscamos cuestionar algunos modelos y aplicaciones de aprendizaje automático y subrayar, a partir de estos ejemplos, la necesidad de identificar y mitigar riesgos asociados al uso de sistemas de Inteligencia Artificial para lograr una práctica responsable, que conecte el desarrollo tecnológico con una sociedad más justa.

Abstract. To the extent that artificial intelligence is gaining ground in different fields of knowledge, permeating everyday life and acquiring the potential to make an impact on our lives, it becomes important to reflect on the social implications of our use of algorithms and the blind spots this could have. This article seeks to question certain models and applications of machine learning and emphasize, through these examples, the need to identify and mitigate the risks associated with the use of artificial intelligence systems to ensure responsible practices that connect technological development with a more just society.

Marisol Flores Garrido es profesora de tiempo completo en la ENES-Morelia. Trabaja en el área de aprendizaje automático y tiene un interés particular en la intersección entre las ciencias computacionales y las humanidades.

Recibido el 5 de julio de 2019; aceptado el 1 de agosto de 2019.