
Entrevistas

“Cuando los algoritmos se apoderan de la administración y políticas públicas, el potencial de daño es ilimitado”. Entrevista con Cathy O’Neil



“The potential for harm when algorithms take over bureaucracies and policies is unbounded”. Interview with Cathy O’Neil

“Quando os algoritmos assumem o controle da administração e das políticas públicas, o potencial de danos é ilimitado”. Entrevista com Cathy O’Neil

 Santiago Liaudat

Laboratorio de Estudios en Cultura y Sociedad, Facultad de Trabajo Social, Universidad Nacional de La Plata, Argentina

 Lucía Céspedes

Érudit, École de bibliothéconomie et des sciences de l’information, Université de Montréal, Canadá

Ciencia, Tecnología y Política

vol. 7, núm. 13, 2024

Universidad Nacional de La Plata, Argentina

ISSN: 2618-2483

ISSN-E: 2618-3188

Periodicidad: Semestral

revista.ctyp@presi.unlp.edu.ar

Recepción: 05 junio 2024

Aprobación: 24 septiembre 2024

DOI: <https://doi.org/10.24215/26183188e120>

URL: <https://portal.amelica.org/ameli/journal/214/2145144002/>

Resumen: Entrevista con Catherine (“Cathy”) Helen O’Neil, nacida en los Estados Unidos, conocida mundialmente por sus estudios críticos sobre los efectos negativos de los algoritmos. Doctora en Matemática por la Universidad de Harvard, es autora de los libros *Haciendo Ciencia de Datos* (2013), *Armas de Destrucción Matemática* (2016) y *La máquina de la vergüenza* (2022). Luego de trabajar para el sector financiero como científica de datos, puso en marcha ORCAA, una empresa de auditoría algorítmica. Es colaboradora habitual de la agencia de noticias Bloomberg Opinion, autora del blog <mathbabe.org> y miembro del Laboratorio Tecnológico de Interés Público de la Escuela de Gobierno John F. Kennedy de la Universidad de Harvard.

Palabras clave: inteligencia artificial, sesgos algorítmicos, discriminación, regulaciones, ética.

Abstract: Interview with Catherine (“Cathy”) Helen O’Neil, born in the United States, known worldwide for her critical studies on the negative effects of algorithms. PhD in Mathematics from Harvard University, she is the author of the books *Doing Data Science* (2013), *Weapons of Math Destruction* (2016) and *The Shame Machine* (2022). After working for the financial sector as a data scientist, she launched ORCAA, an algorithmic auditing firm. She is a regular contributor to Bloomberg Opinion news agency, author of the blog <mathbabe.org> and member of the Public Interest Tech Lab at the Harvard Kennedy School.

Keywords: artificial intelligence, algorithmic bias, discrimination, regulations, ethics.

Resumo: Entrevista com Catherine (“Cathy”) Helen O’Neil, nascida nos EUA, conhecida mundialmente por seus estudos críticos sobre os efeitos negativos dos algoritmos. Com doutorado em Matemática pela Universidade de Harvard, ela é autora dos livros *Doing Data Science* (2013), *Weapons of Mathematical Destruction* (2016) e *The Shame Machine* (2022). Depois de trabalhar no setor financeiro como cientista de dados, ela fundou a ORCAA, uma empresa de auditoria algorítmica. É colaboradora regular da agência de notícias Bloomberg Opinion, autora do blog <mathbabe.org> e membro do Laboratório de Tecnologia de Interesse Público da Escola de Governo John F. Kennedy da Universidade de Harvard.

Palavras-chave: inteligência artificial, vieses algorítmicos, discriminação, regulamentações, ética.

Las críticas a la expansión de las plataformas digitales se han dirigido sobre todo a la pérdida de la privacidad. Pero en su libro *Armas de destrucción matemática* usted fue más allá al afirmar que los algoritmos utilizados para la toma de decisiones ahondan las desigualdades sociales. ¿Qué rasgos hacen de un modelo matemático un arma de destrucción masiva? ¹

Un arma de destrucción matemática es un algoritmo que mucha gente utiliza para tomar decisiones importantes, ya sea sobre su trabajo, sus finanzas, sus libertades o su información. Se lo puede pensar como un sistema de calificación, o tal vez múltiples puntuaciones en distintas dimensiones. Y es secreto. La gente no lo entiende, a menudo ni siquiera saben que se les está puntuando. Y también es injusto. Ningún algoritmo es perfecto, por lo que siempre va a ser injusto con alguien, especialmente si está tomando decisiones sobre quién merece algo. Pero no sólo es injusto de forma arbitraria. Es injusto sistemáticamente, por lo que, de manera habitual, va en contra de los pobres, de las comunidades marginadas, etcétera. La premisa básica de la ciencia de datos es averiguar quién tiene dinero y darle oportunidades, y averiguar quién no lo tiene y aprovecharse de él. Esa es la forma en que Internet funciona como un modelo de negocios. Eso es lo que están haciendo. La preocupación por la privacidad siempre me ha molestado. No es que no crea que merezcamos privacidad, sino que es demasiado evidente que no la tenemos. La gente no termina de entender lo malo que es, porque puedes ser extremadamente bueno en mantener tu privacidad, pero en el momento en que tienes que interactuar con una burocracia, tienes que dar tu número de Seguridad Social, tienes que dar tu dirección, tienes que dar la información que te piden, y te van a pedir la información que les permita perfilarte basándose en la información recolectada de todos los demás usuarios que no tienen el mismo cuidado.² Ese es un punto importante que creo que la gente que está interesada en la privacidad pasa por alto: la privacidad sólo dura hasta que hay que interactuar con la máquina, y entonces desaparece. Y la idea de que nunca tienes que interactuar con una burocracia en toda tu vida es ilusoria. Así que realmente no hay protección para nadie cuando funciona así.

¿Cuáles son algunas de las desigualdades que producen y reproducen los algoritmos? ¿Perpetúan de algún modo una mirada colonialista sobre los países del Sur Global?

A grandes rasgos, mi trabajo como científica de datos consistía en hacer que la gente con suerte tuviera más suerte, y la gente con mala suerte tuviera aún menos suerte. Por supuesto, los algoritmos hacen más que eso, pero sin duda el criterio más fuerte es la riqueza. Después, el género, y después, la raza. Pero la riqueza es lo más fácil de encontrar. No soy experta en colonialismo, pero diría que lo reproduce, absolutamente, porque los algoritmos están completamente centrados en lo que piensa Elon Musk. Se centran en la gente blanca, se centran en la gente occidental, se centran en los estadounidenses. Esa es la perspectiva de ChatGPT y todos los productos que están saliendo de allí. Por lo tanto, en la medida en que todo el mundo está consumiendo el contenido que se construye en California, la respuesta es que absolutamente sí.

¿Cree que afectan también a la democracia? ¿Existe alguna relación entre el modelo de sociedad subyacente a estas armas de destrucción matemática y el ascenso de una nueva derecha en Occidente?

Definitivamente, creo que sí. El subtítulo de mi libro es “Cómo el big data aumenta la desigualdad y amenaza la democracia”. Así que esa es mi hipótesis. Escribí ese libro en 2014, siendo miembro de Occupy en ese momento.³ Luego pude agregar un par de frases en 2016, cuando las elecciones estaban en marcha. Por ejemplo, me las arreglé para agregar algo sobre Cambridge Analytica⁴ y sobre cómo los políticos ya no tienen que decir lo mismo a todo el mundo: pueden enviar mensajes diferentes a personas diferentes. Todo eso, por supuesto, es cierto, pero lo que realmente no anticipé —que para mí, de alguna manera, es aún peor— es que los políticos ni siquiera tienen que tener información, ya no tienen que tener plataformas, solo necesitan ejercer manipulación emocional. Ese es un poco el mensaje, esa es su plataforma, asustarte de esta manera. Acerca de la inmigración, por ejemplo, simplemente hacerte sentir miedo. Esa no es una plataforma típica, pero termina siendo como realmente se ven los mensajes cuando se habla de *microtargeting* en Internet. Está tan hecho a medida, es tan antidemocrático. Los mensajes son a menudo antidemocráticos porque intentan socavar la confianza de la gente en las elecciones. El ecosistema publicitario de Facebook, basado en algoritmos, es una perfecta máquina de propaganda antidemocrática.



Stiftelsen - Internetdagarna 2017

[WikiMedia Commons](#)

El rápido desarrollo de la Inteligencia Artificial (IA) en los últimos años, ¿ha agravado el panorama que describió en 2016?

No puedo creer todo el bombo publicitario que se le ha dado a la IA generativa, que me parece exageradísimo. Para mí, todas estas cosas no son intrínsecamente malvadas hasta que se utilizan con fines malvados. A veces me gusta pensar en escenarios en el que un algoritmo maligno podría usarse para el bien y puedo hacerlo, siempre puedo hacerlo. Por ejemplo, con los algoritmos de riesgo,⁵ que creo que son los más malvados que he conocido: condenan a la gente a sentencias más largas, porque gente similar a ellos fue arrestada y reincidió en el pasado. Pero podríamos utilizar estos algoritmos para mirarnos a nosotros mismos, al público, y preguntarnos por qué la gente pobre va tanto más a la cárcel, por qué la gente sin acceso a tratamientos de salud mental pasa por la cárcel, por qué le hacemos esto a la gente. Escribí el libro para pedir al público que deje de confiar ciegamente en los algoritmos y el *big data*, como se llamaba entonces... ahora se llama IA. Es un nuevo término de marketing, pero la confianza ciega es, una vez más, lo que me preocupa. No me preocupa que estas herramientas sean muy buenas, porque no lo son. En realidad quiero escribir un ensayo sobre esto porque quiero dejar bien claro que es intencional. Es un diseño intencional de estos tipos blancos y ricos de Silicon Valley para que la gente confíe en el chat. De hecho, estaba leyendo sobre retórica en la antigua Grecia y cómo los griegos pensaban sobre argumentos persuasivos, y hay un proceso de seis pasos y uno de ellos es admitir tus errores. Ahí me di cuenta, es como cuando le dices a ChatGPT que cometió un error, siempre te responde algo como “oh, lo siento mucho, eso fue un error”. Se disculpa, lo que le da credibilidad. Pero también mencionaré, porque lo hice como prueba, que si lo corriges en algo que hizo bien, ¡también se disculpa! Pero no hacemos esto último porque somos humanos y no somos desconfiados, estamos entregados a confiar, así que sólo corregimos los errores. No corregimos cosas que son correctas. Lo que quiero decir es que está deliberadamente diseñado para que confiemos en él. Para mí, si tengo que sintetizar en una frase lo que estoy tratando de hacer, es que estoy tratando de evitar que la gente confíe en la máquina y eso es lo que realmente me desespera de esta hipérbole actual de la IA, que es un intento deliberado para que la gente confíe en ella, a pesar de que sigue siendo terrible.

Los problemas detectados en relación con los sesgos y la discriminación por motivos algorítmicos, ¿son intrínsecos a los algoritmos o meras cuestiones de diseño? La ciencia y la tecnología implicadas, ¿son un instrumento neutro cuyos efectos dependen de quién las utilice? ¿O es necesario rediseñarlas?

Es un poco de las dos cosas, dependiendo de cómo lo veas. También depende de lo que entiendas por “herramienta”. La única empresa del mundo que va a usar el algoritmo de noticias de Facebook es Facebook, ¿no? Y lo construyeron para mantener a la gente en Facebook. Es decir, está optimizado para generar interacción. Esa fue una elección porque la interacción se traduce en lucro, por lo que es, literalmente, una herramienta capitalista. Es difícil imaginar un grupo de la sociedad civil diciendo “oh, podríamos utilizar esa herramienta para el bien”, porque no pueden, porque está construida para maximizar el lucro e ignorar cualquier tipo de daño colateral. Está construida de esa manera. Nadie querría eso, excepto Facebook. Pero si se piensa en la herramienta a nivel de los propios algoritmos, o de los canales de información, entonces es fácil imaginar grupos de la sociedad civil que digan: “creo que podríamos construir un canal de información mucho mejor que realmente proporcione información útil, buena y verdadera a la gente”. Así que claro que

vas a pensar “por supuesto que podría hacerse mucho mejor”. Supongo que la respuesta corta es que, dado que estos sistemas algorítmicos son masivos, difíciles de mantener y extremadamente caros, realmente sólo los poseerán grupos de personas muy bien financiados y sofisticados desde el punto de vista técnico. Hasta que esto cambie, serán básicamente las grandes empresas con fines de lucro las que los utilicen para ganar más dinero. Así que ahora mismo es una herramienta de los capitalistas para ser capitalistas. Eso no sería un imperativo en principio, pero es empíricamente lo que está sucediendo.

¿Puede el libre mercado autorregularse a este respecto o deben establecerse normativas estatales?

Jamás diría que podemos confiar en un mercado libre para resolver cualquiera de estos problemas. Creo que los algoritmos están sustituyendo a toda burocracia a la hora de decidir quién merece un empleo, quién merece una tarjeta de crédito, quién merece la cárcel. Cada proceso burocrático de toma de decisiones se está convirtiendo en un algoritmo, si es que ya no lo ha hecho. Incluso se puede argumentar que las propias políticas se están convirtiendo en algoritmos. Las políticas y las decisiones políticas se están “algoritmizando”. Yo lo veo así. Eso significa que estos procesos burocráticos de toma de decisiones están siendo propiedad de los capitalistas para decidir el destino de las personas. Obviamente, como ciudadanos tenemos interés en asegurarnos de que se haga de forma justa, y es absolutamente necesario que lo hagamos. Ahora mismo tenemos un montón de leyes contra la discriminación en Estados Unidos que se ignoran porque no saben cómo aplicarlas a un algoritmo. Por eso fundé mi empresa y llevo años trabajando en ello. Hay progreso, es lento, pero hay progreso en el ámbito de la contratación, hay progreso en el ámbito de los seguros. Cuando los algoritmos se apoderan de la administración y política públicas, el potencial de daño que tienen es ilimitado, así que, por supuesto, tenemos que imponer normas para controlar y mitigar ese daño.

¿Qué lugar ocupan las iniciativas privadas de auditoría y rendición de cuentas en la gobernanza de los sistemas algorítmicos? ¿Cuál es la experiencia de ORCAA al respecto?⁶

Fue muy reveladora. Ahora soy un poco más cínica de lo que era. El progreso es lento. La mayoría de las empresas no quieren ser auditadas. No van a ofrecerse voluntariamente. Quieren seguir teniendo la posibilidad de negar los daños que sus algoritmos están generando. Así que la pregunta es, ¿cómo podemos influir? ¿Cómo podemos hacer que las empresas que realmente queremos auditar necesiten una auditoría? O pongámoslo de esta manera, que ser auditadas les resulte menos costoso que no serlo. Eso es realmente lo que buscamos. Ahora estamos trabajando para varias agencias de aplicación como el Fiscal General y la Comisión Federal de Comercio y otras agencias federales y comisionados de seguros. Pero también estamos trabajando con grupos de abogados que están llevando adelante acciones colectivas, presentando demandas en nombre de grandes grupos de personas, lo que nos va a conseguir influencia, porque así se hace en los Estados Unidos. Se necesita influencia, ya sea de la presión reguladora o del riesgo de litigio, porque la mayoría de las empresas no creen que vayan a tener problemas por el daño que está causando su IA. No les gusta pagar por Recursos Humanos (RRHH), así que despiden a todo su personal de RRHH, lo sustituyen por un algoritmo que es racista, sexista y discriminatorio por edad, y dicen “sí, nadie se mete en problemas por eso, así que ganamos”.

¿Cómo podrían incluirse criterios éticos en el diseño de un algoritmo? ¿A qué se refiere cuando dice que “la ética de la IA no puede automatizarse”?

Los algoritmos no son más que burocracias codificadas. Por eso deben ponerse a prueba, deben diseñarse muy deliberadamente teniendo eso en cuenta. Tienen que equilibrar los derechos de las diferentes partes interesadas y los perjuicios de los distintos grupos interesados. Un grupo puede preocuparse por los falsos negativos y otro por los falsos positivos, y otro por algo totalmente distinto. Y hay que gestionar todas esas necesidades, o al menos reconocerlas en el diseño. Ese es el primer marco que utilizamos, llamado Matriz Ética. Consideramos todas las partes interesadas y todas las preocupaciones. Por ejemplo, lo que podría ir mal para estas personas o incluso para el medio ambiente, que podría ser una parte interesada. La razón por la que digo que no se puede automatizar es que se trata de una cuestión muy sensible al contexto. Estás construyendo un algoritmo para un contexto específico que definirá las partes interesadas y sus preocupaciones. Por poner un ejemplo, podemos hablar de reconocimiento facial. Se utiliza durante el día en un lugar de negocios para asegurarse de que las personas que entran en el edificio trabajan allí, y si no les reconoce la cara, tienen que pasar por un guardia de seguridad para que las deje entrar. Ese es un contexto, otro sería que un algoritmo sea utilizado por la policía con capturas de imagen del video de un robo, y luego la gente sea arrestada. Un error en el primer caso es bastante poco importante, un error ahí implica solo una molestia para alguien. Un error en el segundo contexto podría llevar a alguien a la cárcel por dos semanas, que es un gran problema. La cuestión es que no se puede auditar el reconocimiento facial en general, sólo se puede auditar un caso concreto de uso del reconocimiento facial. A eso me refiero cuando digo que no se puede automatizar.

En su último libro, *La máquina de la vergüenza*, hace una lectura social y política de este sentimiento, demostrando que existe un “complejo industrial de la vergüenza”. ¿Podría ampliar este concepto y enfoque?

La vergüenza es realmente importante y útil y, al mismo tiempo, está muy sobreexplotada, sobre todo para golpear a las personas vulnerables. Tampoco es nada nuevo. Es algo muy antiguo que se ha utilizado durante siglos en el contexto del comercio. Si piensas en la vergüenza con fines de lucro, si piensas en las cremas antiarrugas, te das cuenta de que esto no es nada nuevo. Entonces, ¿qué tiene de nuevo? Algo parecido a lo que decimos del *microtargeting* en política. El ecosistema de la publicidad permite a los anunciantes microfocalizar y avergonzar a la gente con tanta precisión que se ha convertido en un sistema enormemente selectivo y brutal, y ha crecido masivamente. El modelo lucrativo de la vergüenza es, en primer lugar, asegurarse de que alguien se siente avergonzado por algo, y luego venderle un producto que en realidad no resuelve ese problema, para tener un cliente permanente. Hay un artículo que realmente me metió en este tema, así como mi propia investigación sobre la cirugía bariátrica.⁷ El artículo trataba sobre el hecho de que las empresas de máquinas de afeitar querían vender equipos de afeitado a las mujeres asiáticas, pero las mujeres asiáticas no se sentían mal por tener pelos en las piernas. Así que primero hicieron una campaña para que las mujeres asiáticas se sintieran avergonzadas de su vello corporal, y luego les vendieron las maquinillas de afeitar. Y funcionó. Así es como funciona. Tienes que asegurarte de que la gente se sienta avergonzada.

¿Qué ejemplos puede dar de este uso industrial de la vergüenza para humillar? ¿Quién se beneficia de golpear a las personas vulnerables?

Las grandes empresas tecnológicas, las redes sociales, porque son las que nos han entrenado deliberadamente, sin que nos demos cuenta, para avergonzarnos unos a otros. Me gusta el ejemplo del producto OMV!, dirigido a las adolescentes para que se avergüencen del olor de su vagina.⁸ Luego, se les vende un producto que no las ayudará, ya sea un producto dietético o un producto para el cuidado de la piel... o cualquier otra cosa relacionada con el cuerpo. Son productos que no están diseñados para ayudarte, solo para aliviar tu vergüenza temporalmente. Hasta que ya no lo hace. Como dije, no es nuevo, pero está increíblemente bien orquestado en Internet. Ahora también se avergüenza a los varones por sus cuerpos. Es un gran tema empezar a avergonzar a los jóvenes desde temprano, porque van a ser clientes de por vida de estos productos. Precisamente, la otra cosa de la que trata mi libro es de cómo ya no son sólo estas empresas: somos nosotros. Las redes sociales nos han cooptado en ese sistema de humillación. Si las empresas de cuidado de la piel son la vieja guardia de la vergüenza con ánimo de lucro, las redes sociales son la nueva guardia. No nos avergüenzan deliberadamente para luego obligarnos a comprarles algo, sino que crean una plataforma para que nos avergoncemos unos a otros y ganan dinero con esa interacción, con esas luchas que mantenemos todos los días en Internet. Pienso en ello como un resultado directo de lo rentable que es la vergüenza.

En su libro también señala una relación entre la pseudociencia y el complejo industrial de la vergüenza. ¿Podría decirnos cuál es ese vínculo?

Vengo pensando en eso recientemente, es un tema que surge una y otra vez. Creo que es porque soy matemática de formación, y no tengo esa reacción tan común y tan explotada de sentir que “esto es ciencia, así que tengo que confiar en ello”. Antes hablábamos de la confianza ciega en la IA. Esa es mi lucha, pero me he dado cuenta de que la pelea que tengo una y otra vez, siempre es con la pseudociencia de una forma u otra. Es el hecho de que la gente se siente intimidada por la jerga pseudocientífica o los argumentos pseudomatemáticos. Me encuentro en esta posición afortunada en la que digo “¿por qué estás confiando en esto?”. Pero en cuanto la gente confía en algo, le concede autoridad a esa cosa. Esa es una configuración perfecta para la vergüenza, porque en última instancia, si alguien que no tiene ninguna autoridad sobre ti trata de avergonzarte, vas a pensar “pero por favor, fuera de aquí”. Pero si alguien con autoridad te avergüenza, entonces eso duele. Así que es importante ganar autoridad para avergonzar a otros y vender ese producto. Por eso creo que hay una conexión tan fuerte entre la pseudociencia y el complejo industrial de la vergüenza. Necesita afirmarse en la autoridad, en un cierto tipo de experiencia, y luego apelar a las emociones de alguien.

La máquina de la vergüenza produce ganadores y perdedores, y culpa a estos últimos de sus fracasos. Esto puede interpretarse como una forma de trasladar la responsabilidad de los problemas sociales de las instituciones a los individuos. ¿Es el complejo industrial de la vergüenza un instrumento del neoliberalismo?

Totalmente. Creo que uno de los mejores ejemplos podría ser el ahorro para la jubilación. Hace cincuenta años teníamos sistemas de ahorro para la jubilación que eran bastante buenos, al menos para los blancos de clase media, con pensiones, sobre todo si estabas afiliado a un sindicato. La seguridad social también era bastante buena. Hemos estado recortando la seguridad social y los neoliberales han barrido con las pensiones. Ahora casi nadie tiene pensión y se espera que todo el mundo ahorre por sus propios medios para la jubilación. Pero nadie sabe cuánto tiempo va a estar jubilado porque nadie sabe cuándo va a morir. Así que todo el mundo tiene que ahorrar teóricamente una cantidad infinita de dinero, mucho más dinero del que

probablemente necesitará en realidad, por lo que en realidad es muy ineficiente económicamente. Pero este sistema se deshizo de un asunto público ya acordado, y lo convirtió en responsabilidad individual. Y no está funcionando. La gente no tiene dinero para la jubilación, simplemente no está funcionando. Pero en lugar de reconocer que fue una mala idea deshacerse de las pensiones y recortar la seguridad social, lo que se dice en la calle es que esa gente debería avergonzarse de no haber ahorrado porque “¿qué estaban haciendo? Fueron tan descuidados”. Así que es un gran ejemplo de cómo se empieza con un problema social reconocido, luego se lo individualiza, luego se culpa a la gente por no resolverlo individualmente y luego simplemente tienen que sufrir.

Como experta en datos y referente internacional en la crítica al mal uso de algoritmos y tecnologías digitales, ¿qué futuro vislumbra como posible y qué acciones podemos emprender para construir un futuro deseable bajo el actual orden tecnológico?

Tengo una visión pesimista y otra optimista. Mi visión pesimista me abrumó la semana pasada cuando busqué algo en Google y, en lugar de obtener enlaces, obtuve una mala respuesta que no tenía ningún sentido, y entonces me imaginé un futuro en el que eso fuera todo lo que pudiéramos hacer. Nos habríamos deshecho de nuestros libros, sólo tendríamos Internet, y si tratáramos de hacer una investigación de fondo sobre nuestras preguntas sólo obtendríamos más respuestas de inteligencia artificial. Estaríamos recibiendo literalmente los balbuceos inexactos y pre-digeridos de personas como los usuarios de Reddit y es todo cuesta abajo, ya no habría manera de verificar realmente los hechos.⁹ Esa es una visión. La pregunta es cuánta gente se opondría a eso y cuál sería nuestro poder para detenerlo. En particular, teniendo en cuenta lo disfuncional que es el Congreso para cambiar cualquier cosa, especialmente las cosas con las que algunos se benefician de cierta manera. Así que esa es una visión muy pesimista en la que básicamente nos están dando lo que las empresas tecnológicas deciden que es lo suficientemente bueno, en su opinión, para nosotros. Pero, con optimismo, no creo que eso sea realmente tan probable. Creo, con optimismo, que hay ciclos de hipérbole. Son estas olas de confianza ciega. Creo que ésta va a remitir porque es evidente que la IA no es muy inteligente y no tiene ningún modelo de verdad y no puede considerarse sabia en absoluto. La cuestión es si podemos imponer el escepticismo a estos sistemas en lugar de aceptarlos sin más. La otra cosa optimista que siento es que incluso durante el gobierno de Trump —que obviamente no tenía ningún afecto por mí y mi trabajo— me las arreglé para hacer mucho con los individuos. Cuando digo “hacer mucho” me refiero a pruebas de concepto en la empresa en el contexto de los seguros o lo que sea, diciendo “fíjate este análisis del seguro del automóvil y cómo está cobrando a los conductores negros mucho más que a los conductores blancos, aquí hay un análisis y así es como lo arreglamos, y así es como cumplimos la ley contra la discriminación”. Es posible hacerlo. Es posible tomar una sentencia ética en inglés llano y traducirla en una regla, en código, y hacerle un seguimiento y decir “este algoritmo está siguiendo esta regla que nos importa” o “este algoritmo no está siguiendo, esto está fallando por este margen”. Hay una manera de abordarlo en cualquier contexto particular y una vez que tengamos suficientes ejemplos, lo ideal es que se sepa que realmente podemos asegurarnos de que los algoritmos no son sólo destructivos, sino que trabajen para la gente, no sólo para las personas que los poseen. Tampoco sé hasta qué punto eso es probable, creo que es algo intermedio.

¿Conoce alguna normativa en el mundo que se esté aplicando para un uso más ético, más justo, de los algoritmos?

Lo más avanzado que he oído es la Ley de Inteligencia Artificial de la Unión Europea.¹⁰ Todavía no se ha convertido en una política efectiva. Por ahora sólo se basa en principios, tiene que traducirse en lo que realmente significa, pero está haciendo muchas cosas bien. Por ejemplo, clasifica los algoritmos según su importancia, su impacto y su potencial nocivo en determinados casos de uso. Así que se preocupa por el contexto, se preocupa por los resultados en los seres humanos, y tiene un escrutinio mucho mayor para los algoritmos de alto impacto, potencialmente dañinos. Es lo mejor que he visto. Por supuesto, hablo todo el tiempo con los responsables políticos de Estados Unidos y algunas de las cosas que están saliendo son buenas. La pregunta es: ¿se van a aprobar? Y la segunda pregunta es: ¿se van a aplicar? Pero podría ser. Sin dudas, es mucho más complicado. Pero diré esto: cuando escribí ese libro en 2014, no pude encontrar a nadie con quien hablar del tema. A nadie. Y ahora hay conferencias y hablo con gente todo el tiempo cuyo trabajo es pensar en política e IA. Así que está muy arriba en la agenda y eso es emocionante. Pero aún no hemos visto nada bueno.

Usted mencionaba que esto forma parte de la agenda que debaten los responsables políticos. ¿Cuál cree que debe ser el papel de los científicos en este debate?

Es muy difícil. No lo sé. Ya no soy académica. Pero si quisiera decir algo un poco simplista, diría que lo más importante es dejar de actuar como si las STEM [Ciencia, Tecnología, Ingeniería, Matemáticas] fueran lo único que importa y empezar a hacer hincapié en las artes y la poesía. Ese es uno de nuestros problemas, que elevamos la ciencia, las matemáticas y la tecnología como si fueran más importantes que el amor. Es ridículo.

NOTAS

- 1 En inglés se produce un juego de palabras que en español se pierde entre “armas de destrucción masiva” (weapons of mass destruction) y “armas de destrucción matemática” (weapons of math destruction).
- 2 En todo el texto, se debe tener en cuenta que la palabra “burocracia” en inglés no tiene un sesgo peyorativo tan marcado como en español, donde, al menos en Argentina, se asocia a procedimientos poco ágiles. El diccionario de la Real Academia Española recoge esta doble acepción en el español: “1. Organización regulada por normas que establecen un orden racional para distribuir y gestionar los asuntos que le son propios. 2. Influencia excesiva de los funcionarios en los asuntos públicos. 3. Administración ineficiente a causa del papeleo, la rigidez y las formalidades superfluas”. El uso que hace O’Neil del término remite al primer uso identificado por la RAE.
- 3 Occupy Wall Street fue un movimiento de izquierda contra la desigualdad económica, la avaricia empresarial, las grandes finanzas y la influencia del dinero en la política que comenzó en 2011 en el Parque Zuccotti, situado en el distrito financiero de Nueva York, y se expandió a otras grandes ciudades de Estados Unidos y del mundo.
- 4 Cambridge Analytica fue una consultora política británica que utilizó datos recopilados de millones de usuarios, sin su consentimiento, a través de una aplicación de Facebook, para proporcionar asistencia analítica a las campañas presidenciales de Ted Cruz y Donald Trump en los Estados Unidos, Mauricio Macri en Argentina, entre otras figuras políticas.
- 5 Los algoritmos de riesgo hacen predicciones sobre el riesgo delictivo de una persona a partir de los datos y estadísticas introducidos en el sistema. Las metodologías utilizadas por el software para predecir el riesgo son un secreto comercial, y las máquinas sólo están obligadas a comunicar sus estimaciones de reincidencia a los jueces.
- 6 O’Neil Risk Consulting & Algorithmic Auditing (ORCAA), empresa fundada en 2016 por Cathy O’Neil, que ofrece auditoría, consultoría y formación sobre sistemas algorítmicos e inteligencia artificial. Ver <https://orcaarisk.com/>
- 7 La cirugía bariátrica es el conjunto de procedimientos quirúrgicos usados para tratar la obesidad, buscando la disminución del peso corporal.
- 8 OMV! es un producto comercial de la empresa estadounidense Vagisil, lanzado en 2021, muy cuestionado por ginecólogos y expertos en salud femenina. Véase <https://www.nytimes.com/2021/02/18/well/vagisil-omv-teens.html>
- 9 Reddit es un sitio web de marcadores sociales y agregador de noticias donde los usuarios pueden añadir textos, imágenes, videos o enlaces. Los usuarios pueden votar a favor o en contra del contenido, haciendo que aparezcan en las publicaciones destacadas. Reddit está entre los diez sitios web más visitados del mundo.
- 10 Ver: <https://artificialintelligenceact.eu>

AmeliCA

Disponible en:

<https://portal.amelica.org/ameli/ameli/journal/214/2145144002/2145144002.pdf>

Cómo citar el artículo

Número completo

Más información del artículo

Página de la revista en portal.amelica.org

AmeliCA
Ciencia Abierta para el Bien Común

Santiago Liaudat, Lucía Céspedes

“Cuando los algoritmos se apoderan de la administración y políticas públicas, el potencial de daño es ilimitado”.

Entrevista con Cathy O’Neil

“The potential for harm when algorithms take over bureaucracies and policies is unbounded”. Interview with Cathy O’Neil

“Quando os algoritmos assumem o controle da administração e das políticas públicas, o potencial de danos é ilimitado”. Entrevista com Cathy O’Neil

Ciencia, Tecnología y Política

vol. 7, núm. 13, 2024

Universidad Nacional de La Plata, Argentina

revista.ctyp@presi.unlp.edu.ar

ISSN: 2618-2483

ISSN-E: 2618-3188

DOI: <https://doi.org/10.24215/26183188e120>