# MINIMIZACIÓN DE DATOS

## FUNDAMENTAL PARA LA PROTECCIÓN DE LA PRIVACIDAD Y LA REDUCCIÓN DE DAÑOS

accessnow

Access Now defiende y extiende los derechos digitales de los usuarios en riesgo alrededor del mundo. Mediante la combinación de apoyo técnico directo, campañas globales, el análisis integral de políticas públicas, el financiamiento a grupos locales emergentes, intervenciones judiciales y eventos como RightsCon, luchamos por los derechos humanos en la era digital.

# **MINIMIZACIÓN DE DATOS:**
FUNDAMENTAL PARA LA PROTECCIÓN DE LA PRIVACIDAD Y LA REDUCCIÓN DE DAÑOS

Este documento es una publicación de Access Now. Fue escrito por Eric Null, Isedua Oribhabor y Willmary Escoto. Agradecemos a los y las miembros de Access Now que nos brindaron su apoyo, incluyendo a Estelle Massé, Daniel Leufer, Gaspar Pisanu, Jennifer Brody, Sage Cheng, Juliana Castro y Donna Wentworth.

**MAYO DE 2021**

# ÍNDICE

# INTRODUCCIÓN A LA MINIMIZACIÓN DE DATOS: QUÉ ES Y POR QUÉ ES IMPORTANTE

Si bien la minimización de datos no tiene una definición única, la más sencilla y útil de las definiciones es que las organizaciones (ya sean empresas privadas, entidades públicas u organismos gubernamentales) que recopilen datos deben recopilar solo aquellos que sean necesarios para brindar su producto o servicio. La minimización de datos hace referencia a recopilar datos solo para un fin inmediato y necesario, sin acumular datos por la "remota posibilidad de que sean útiles en el futuro".[1] Específicamente, las organizaciones deberían limitar: 1) la extensión de los datos que recopilan; 2) la cantidad de datos que recopilan dentro de esa acotada extensión; y 3) la retención de dichos datos.[2] Un ejemplo claro del principio de minimización de datos en la práctica es que "un controlador de datos no procesará de manera constante la ubicación precisa y detallada de un vehículo con el propósito de mantenimiento técnico u optimización del modelo".[3]

La minimización de datos es un principio central en la protección de datos personales y es parte de los "principios prácticos de información justa".[4] También es parte de un concepto dominante al que se refiere como "privacidad por diseño y por defecto", que alienta a las organizaciones a incorporar la privacidad en sus productos y servicios de antemano, en lugar de pensar en ella como algo secundario. Otros principios de privacidad por diseño incluyen la retención de datos limitada y la limitación del propósito. La minimización está interconectada con estos diferentes conceptos debido a que, dado que los datos ya no son necesarios para cumplir el propósito inmediato y necesario, las organizaciones ya no deberían retenerlos.

Por lo general, las organizaciones no se toman en serio la minimización de datos. Recopilan y retienen datos impunemente, y muchas de ellas no brindan salvaguardas ni toman un enfoque disciplinado para proteger la privacidad individual. Un estudio de empresas de Europa señala que el 72 % de los

[1] International Commissioner's Office. *Principio (c): Minimización de datos,* https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation.

[2] Primero, "debería minimizarse la *posibilidad* de recopilar datos personales sobre terceros". Luego, "dentro de las posibilidades restantes, debería minimizarse la *recopilación* datos personales". Finalmente, "debería minimizarse el *tiempo* durante el cual se almacenarán los datos personales recopilados". *Terminology for Talking about Data Minimization*, IETF (2010), https://tools.ietf.org/id/draft-hansen-privacy-terminology-00.html (sin cursivas en el original). La retención de datos es un principio estrechamente relacionado que asegura que, una vez que los datos cumplan su propósito, la organización deba eliminar dichos datos.

[3] Commission Nationale Informatique & Libertes (NCIL), *Compliance Package - Connected Vehicles and Personal Data* (octubre de 2017), https://www.cnil.fr/sites/default/files/atoms/files/cnil_pack_vehicules_connectes_gb.pdf en la página 10.

[4] *Fair Information Practice Principles (FIPP)*, International Association of Privacy Professionals, https://iapp.org/resources/article/fair-information-practices. Tenga en cuenta, de todas formas, que los FIPP definen las limitaciones de recopilación de manera más amplia: "Debe haber límites para la recopilación de datos personales, y todos esos datos deberían obtenerse por medios lícitos y justos y, cuando corresponda, con el conocimiento o consentimiento del sujeto titular de los datos". *Id*. Nuestra definición es más estricta: se centra en los datos que son necesarios para brindar el producto o servicio.

datos recopilados no se utiliza.[5] Otro informe a nivel mundial indica que el 55 % de todos los datos recopilados son "datos oscuros" que no se usan para ningún fin tras ser recopilados.[6] El razonamiento es claro. Las organizaciones quieren tener tantos datos como sea posible para comercializar (por ejemplo, a través de la publicidad conductual), rastrear personas, o hacer uso de los datos en el futuro, potencialmente para entrenar un modelo de *machine learning* o para vender la información a corredores de datos o gobiernos. Los datos se han convertido en una mercancía para muchas organizaciones, y pocas de ellas cambiarán su modelo de negocio para reconocer la privacidad como un derecho humano.[7]

Los desarrolladores de aplicaciones, por ejemplo, han tenido dificultades en torno a la minimización de datos. En 2013, se estudiaron 100 aplicaciones y se descubrió que 56 de ellas, incluidas "Angry Birds" y la infame "Brightest Light", recopilaban información de geolocalización que no les correspondía.[8] La Comisión Federal de Comercio (FTC) de EE. UU. tomó medidas drásticas contra los creadores de las aplicaciones y determinó que las aplicaciones engañaban a las personas y, una vez que las descargaban, estas podían acceder a sus calendarios, su ubicación y la configuración de la cámara.[9]

Más recientemente, el juego móvil "Pokémon GO" de realidad aumentada tampoco minimizó la recopilación de datos. Millones de personas en todo el mundo descargaron la aplicación, lo cual no solo dio acceso a su ubicación y a su cámara (relativamente necesarias para el funcionamiento del juego), sino también a sus cuentas de Google. Esto permitió que Pokémon GO tuviera acceso a sus fotos, calendarios, correos electrónicos y otros documentos, lo cual la crítica catalogó como "un gran riesgo para la seguridad".[10] Luego de que defensores y defensoras de la privacidad criticaran la

[5] *Big Data's Failure: The struggles businesses face in accessing the information they need,* Pure Storage (julio de 2015), https://info.purestorage.com/rs/225-USM-292/images/Big%20Data%27s%20Big%20Failure_UK%281%29.pdf?aliId=64921319.
[6] Companies Collect a Lot of Data, But How Much Do They Actually Use?, Priceonomics, https://priceonomics.com/companies-collect-a-lot-of-data-but-how-much-do.
[7] *Consulte* Eric Null, *Ask Apple: Facebook Doesn't Give a Damn about Privacy Protections*, Access Now (29 de marzo de 2021), https://www.accessnow.org/facebook-apple-privacy-war (donde se explica que Facebook y su CEO, Mark Zuckerberg, estaban furiosos porque Apple había implementado una actualización proprivacidad en el sistema operativo iOS y esta afectaba a Facebook y empresas pequeñas).
[8] Bob Sullivan, *A shock in the dark: Flashlight app tracks your location*, NBC News (16 de enero de 2013), https://www.nbcnews.com/technolog/shock-dark-flashlight-app-tracks-your-location-1B7991120.
[9] Comisión Federal de Comercio, comunicado de prensa, *Android Flashlight App Developer Settles FTC Charges It Deceived Consumers* (5 de diciembre de 2013), https://www.ftc.gov/news-events/press-releases/2013/12/android-flashlight-app-developer-settles-ftc-charges-it-deceived; Robert McMillan, *The Hidden Privacy Threat of … Flashlight Apps?*, Wired (20 de octubre de 2014), https://www.wired.com/2014/10/iphone-apps.
[10] Laura Hudson, *How to Protect Privacy While Using Pokémon Go and Other Apps*, N.Y. Times (12 de julio de 2016), https://www.nytimes.com/2016/07/14/technology/personaltech/how-to-protect-privacy-while-using-pokemon-go-and-other-apps.html.

aplicación por lucrar con los movimientos de quienes la descargaban,[11] Niantic, la empresa desarrolladora de la aplicación, tomó medidas para responder a la reacción pública.[12]

Hay quienes piensan que la recopilación de datos más expansiva (en particular para publicidad conductual) es ventajosa tanto para quienes publican anuncios como para quienes consumen.[13] Sin embargo, hay razones para dudar del alcance de esos beneficios. En primer lugar, un estudio indicó que la mayor parte del valor agregado de la publicidad conductual se la llevan las empresas de tecnología, no las personas ni anunciantes.[14] En segundo lugar, si bien se asume que la publicidad conductual es superior a la contextual, esa presunción no es una verdad absoluta, ya que la publicidad contextual no ha recibido suficientes inversiones en los últimos veinte años.[15] Por último, las personas generalmente consideran siniestro el rastreo en internet por parte de las mismas publicidades, lo que hace que estas pierdan efectividad.[16]

En este documento, analizamos por qué la minimización de datos es fundamental en la protección de la privacidad y por qué puede reducir los daños de la recopilación y explotación de datos personales. Luego, ofrecemos a quienes tomen decisiones, como legisladores y legisladoras, empresas desarrolladoras de software y demás personas involucradas en el desarrollo y la implementación de las políticas de minimización de datos, una guía sobre cómo aplicar los principios de minimización de datos para **abordar los perjuicios relacionados con los derechos civiles, limitar el impacto negativo de la publicidad conductual** y **entrenar sistemas de *machine learning* (ML).**

## POR QUÉ LA MINIMIZACIÓN DE DATOS ES UNA CUESTIÓN DE DERECHOS HUMANOS

La privacidad es un derecho humano y la minimización de datos es una cuestión de derechos humanos. El impacto más importante de una buena minimización de datos es la reducción de daños:

[11] Yehong Zhu, *How Niantic Is Profiting Off Tracking Where You Go While Playing 'Pokémon GO',* Forbes (29 de julio de 2016), https://www.forbes.com/sites/yehongzhu/2016/07/29/how-niantic-is-profiting-off-tracking-where-you-go-while-playing-pokemon-go/#1b6137a56df9.
[12] Nathan Oliverez-Giles, *'Pokemon Go' Creator Closes Privacy Hole but Still Collects User Data*, Wall St. Journal (13 julio de 2016), https://www.wsj.com/articles/pokemon-go-creator-closes-privacy-hole-but-still-collects-user-data-1468363704.
[13] James Ewen, *What Is Behavioral Targeting? - All You Need to Know in 2020*, Tamoco (24 de septiembre de 2019), https://www.tamoco.com/blog/what-is-behavioral-targeting.
[14] Veronica Marotta *et al*., *Online Tracking and Publishers' Revenues: An Empirical Analysis*, taller sobre economía de la seguridad de la información (mayo de 2019), https://weis2019.econinfosec.org/wp-content/uploads/sites/6/2019/05/WEIS_2019_paper_38.pdf.
[15] Becky Chao & Eric Null, *Paying for Our Privacy: What Online Business Models Should Be Off-Limits?,* Open Technology Institute (17 de septiembre de 2019), https://www.newamerica.org/oti/reports/paying-our-privacy-what-online-business-models-should-be-limits (“Como expresó [la asesora jurídica de DuckDuckGo, Megan Gray], la publicidad contextual ha sufrido la falta de inversiones”.).
[16] Leslie K. John *et al*., *Ads That Don't Overstep: How to make sure you don't take personalization too far*, Harvard Business Review Magazine (enero-febrero de 2018), https://hbr.org/2018/01/ads-that-dont-overstep (donde se indica que las personas tienen menos probabilidades de interactuar con publicidades que utilizaron información inferida o información recopilada por terceros).

los datos que no se recopilan no pueden perjudicar a las personas. Si las organizaciones recopilan más datos, crece el daño potencial y real para las personas. Reducir la cantidad de datos que se recopilan es importante por, al menos, dos motivos: las personas no quieren que las organizaciones recopilen cada pieza de información sobre ellas, y la información personal puede utilizarse indebidamente de maneras que perpetúan daños significativos, lo cual sucede a menudo.

## Las personas no quieren que las organizaciones recopilen cantidades excesivas de datos

Generalmente, las personas cuestionan la sensatez de permitir a las organizaciones recopilar cualquier dato sobre ellas. Las personas a menudo no saben cómo las organizaciones usan sus datos y sienten que tienen poco control respecto de las prácticas de datos de las organizaciones.[17] Una encuesta de Pew del 2019 arrojó que el 81 % de las personas encuestadas en EE. UU. sentían que los riesgos potenciales a los que se enfrentan debido a la recopilación de datos superan a los beneficios, y la mayoría (79 %) también informó estar preocupada sobre la manera en que las empresas usan sus datos.[18]

En el resto del mundo, la sensación es la misma. Un estudio que incluyó a más de 25.000 personas de 40 países descubrió que "a siete de cada 10 personas les preocupa compartir su información personal, mientras que a dos tercios de la población mundial no le agradan las actuales prácticas de privacidad de la mayoría de los recopiladores de datos".[19] Otra encuesta basada en Europa halló que "el 41 % [de las personas encuestadas] no quería compartir datos [personales específicos] con empresas privadas".[20] Además, si bien el 72 % conocía la configuración de privacidad de sus teléfonos inteligentes, el 24 % no sabía cómo verificar la configuración de privacidad de las aplicaciones.[21] El 81 % del pueblo australiano considera que es un "uso indebido que una organización solicite información que parece no ser relevante para el propósito de la transacción, un 7 % más que en 2017".[22] Además, mientras que el 85 % de la comunidad australiana encuestada tenía "un buen entendimiento de por qué debería proteger su información personal, [...] el 49 % [dijo] que no sabía cómo hacerlo".[23] Lamentablemente, debido a que la recopilación de datos excesiva es la norma, las

[17] *Better Machine Learning Through Data Minimization,* Privatar (5 de marzo de 2020), https://www.privitar.com/blog/better-machine-learning-through-data-minimization.
[18] Brooke Auxier *et. al., Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information,* Pew Research Center (15 de noviembre de 2019), https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information.
[19] *Global Crisis In Trust Over Personal Data*, Worldwide Independent Network of Market Research (20 de julio de 2020), https://winmr.com/global-crisis-in-trust-over-personal-data.
[20] *Your rights matter: Data protection and privacy - Fundamental Rights Survey,* European Union Agency for Fundamental Rights (18 de junio de 2020), https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-fundamental-rights-survey-data-protection-privacy_en.pdf, en la página 3.
[21] *Id*. en la página 7.
[22] Daniella Kafouri & John Pane, *Australian community attitudes toward privacy survey 2020,* International Association of Privacy Professionals (3 de diciembre de 2020), https://iapp.org/news/a/australian-community-attitudes-towards-privacy-survey-2020.
[23] *Id*.

personas no son capaces de expresar sus preferencias de privacidad, o se les dificulta hacerlo, y terminan resignándose a usar los servicios que desean con prácticas de datos que no les agradan.[24]

## Recopilar cantidades excesivas de datos causa grandes daños

La amplia recopilación de datos ha causado daños significativos, y riesgos de daños, para las personas. Estos daños incluyen desde la usurpación de identidad y daños físicos hasta ejemplos menos obvios, como perjuicios en las relaciones (debido a la pérdida de confidencialidad), daños emocionales o reputacionales (debido a que la información privada se vuelve pública) o efectos inhibidores en el discurso o la actividad (debido a una pérdida de confianza en el gobierno u otras organizaciones).[25]

Los daños relacionados con la discriminación son particularmente preocupantes.[26] La recopilación y el procesamiento de datos pueden reducir oportunidades para la comunidad negra, hispana, indígena y otras comunidades de personas de color, o hacerlas blanco de campañas discriminatorias y engaños.[27] Por ejemplo, durante las elecciones del 2016, la Agencia de Investigación de Internet de Rusia usó la función de filtrado de público de Facebook y Twitter para dirigirse a personas negras para desalentarlas a votar.[28] La investigación arroja pruebas suficientes de discriminación impulsada por datos de la vida diaria, lo que impacta en la vivienda, el empleo, los préstamos, los créditos, y mucho más. Por ejemplo, un estudio llevado a cabo en los EE. UU. descubrió que los sesgos en la "fijación de precios algorítmica estratégica" llevaba a que solicitantes de préstamos de origen afroamericano y latino pagaran tasas de interés más altas en la compra de viviendas y la refinanciación de préstamos en comparación con solicitantes de piel blanca u origen asiático, lo que les cuesta a los solicitantes afroamericanos y latinos entre USD 250 y 500 millones cada año.[29] Otro ejemplo son las empresas que usan filtros de anuncios de Facebook para impedir que ciertos perfiles vean anuncios laborales, como

[24] Joseph Turow, *The Tradeoff Fallacy*, University of Pennsylvania (junio de 2015), https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf ("la mayoría de las personas estadounidenses se resigna a dar sus datos…").

[25] Daniel J. Solove & Danielle Keats Citron, *Privacy Harms*, George Washington School of Law (2021), https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2790&context=faculty_publications.

[26] *Id*.

[27] Cameron F. Kerry, *Federal privacy legislation should protect civil rights*, Brookings Institute (16 de julio de 2020), https://www.brookings.edu/blog/techtank/2020/07/16/federal-privacy-legislation-should-protect-civil-rights.

[28] Scott Shane & Sheera Frenkel, *Russian 2016 influence operation targeted African-Americans on social media,* The New York Times Magazine (17 de diciembre de 2018), https://www.nytimes.com/2018/12/17/us/politics/russia-2016-influence-campaign.html; Jack Stubbs, *Facebook says Russian influence campaign targeted left-wing voters in U.S., UK*, Reuters (15 de septiembre de 2020), https://www.reuters.com/article/usa-election-facebook-russia/facebook-says-russian-influence-campaign-targeted-left-wing-voters-in-u-s-uk-idUSKBN25S5UC; *Report of the Select Committee on Intelligence in the U.S. Senate on Russian Active Measures Campaigns and Interference in the 2016 Election, Volume 2: Russia's Use of Social Media with Additional Views*, en la página 35, https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf (una declaración de Renee DiResta, directora de investigación en New Knowledge, una empresa de ciberseguridad, indicó que "Las narrativas de supresión del votante [...] estaban dirigidas específicamente al público de personas negras".).

[29] Robert Bartlett *et al*., *Consumer Lending Discrimination in the FinTech Era*., Univ. of California School of Law (2017), http://faculty.haas.berkeley.edu/morse/research/papers/discrim.pdf?_ga=2.236934529.1452837941.1619188513-1084532383.1619188513; Laura Counts, *Minority homebuyers face widespread statistical lending discrimination, study finds*, Univ. of California Berkeley Haas School of Business (13 de noviembre de 2018), https://newsroom.haas.berkeley.edu/minority-homebuyers-face-widespread-statistical-lending-discrimination-study-finds.

de Uber y NTB Trucking, que ofrecen empleos solo a varones,[30] y otras 40 empresas que categóricamente excluyen a personas de edad más avanzada para que no visualicen sus anuncios.[31] En la industria de los servicios financieros, diferentes empresas quebrantaron las políticas antidiscriminación de Facebook al dirigir sus anuncios de tarjetas de crédito hasta préstamos hipotecarios a grupos restringidos.[32] Las empresas estadounidenses Staples, Home Depot, Discover Financial Services, y Rosetta Stone usaron datos de ubicación física de los perfiles para mostrar precios más elevados y menores ofertas en línea para personas en barrios de bajos ingresos.[33] Este tipo de datos no debería ser recopilado en primera instancia, salvo que sea necesario para prestar el servicio o, como explicamos más abajo, para auditar los sesgos de los sistemas de procesamiento de datos. Una vez recopilados, los datos no deberían ser usados para discriminar a las personas.

## La minimización de datos reduce daños al limitar la vigilancia y aumentar la seguridad

Otro de los riesgos de la recopilación excesiva de datos es el uso de la información para la vigilancia por parte del gobierno, que puede resultar en abusos por parte de autoridades gubernamentales y causar efectos inhibitorios en la libertad de expresión. La minimización de datos también puede reducir esos daños. Cuando un gobierno busca información de una empresa como Signal, que ofrece comunicaciones cifradas de extremo a extremo (lo que evita que terceros vean el contenido de las comunicaciones) y mantiene al mínimo los datos que recopila de sus usuarios y usuarias, la empresa no tiene información para darles a esas autoridades gubernamentales. Recientemente, cuando el gobierno de los EE. UU. llevó a cabo una solicitud de esa índole, en la que pidió también los nombres y direcciones de las personas, la respuesta de Signal fue que no podía proporcionar tal información. “Es imposible entregar datos a los que nunca hemos tenido acceso para empezar”.[34] Si más empresas adoptaran este tipo de minimización de datos, menos personas serían objeto de violaciones de privacidad, vigilancia gubernamental y abusos.

La minimización de datos también es un elemento importante de la seguridad de los datos. A medida que la innecesaria recopilación y retención de datos aumenta, la creciente mina de datos se convierte en un blanco para terceros, ya sean agencias de aplicación de la ley o hackers maliciosos. El informe de transparencia más reciente de Amazon, que abarca el sitio de compras de Amazon y sus productos Echo, Ring, y Fire, indica un aumento del 800 % en las solicitudes de datos de clientes por parte de

[30] Ariana Tobin & Jeremy B. Merrill, *Facebook Is Letting Job Advertisers Target Only Men*, ProPublica (18 de septiembre de 2018), https://www.propublica.org/article/facebook-is-letting-job-advertisers-target-only-men.
[31] Jeff Larson *et al*., *These Are the Job Ads You Can’t See on Facebook If You’re Older,* ProPublica (19 de diciembre de 2017), https://projects.propublica.org/graphics/facebook-job-ads.
[32] Corin Faife & Alfred Ng*, Credit Card Ads Were Targeted by Age, Violating Facebook’s Anti-Discrimination Policy,* Markup (29 de abril de 2021), https://themarkup.org/citizen-browser/2021/04/29/credit-card-ads-were-targeted-by-age-violating-facebooks-anti-discrimination-policy.
[33] Jennifer Valentino-DeVries *et al*., *Websites Vary Prices, Deals Based on Users' Information*, Wall St. Journal (24 de diciembre de 2012), https://www.wsj.com/articles/SB10001424127887323777204578189391813881534.
[34] *Grand jury subpoena for Signal user data, Central District of California,* Signal (27 de abril de 2021), https://signal.org/bigbrother/central-california-grand-jury.

agencias de aplicación de la ley solo en 2020.[35] Este pico está vinculado a la cantidad de datos que Amazon retiene de su clientela.

El daño causado por las filtraciones de datos, los hackeos o el acceso no autorizado a datos dentro de una organización es demasiado grave para justificar la recopilación de más datos que los necesarios para brindar un producto o servicio. Las organizaciones son responsables de asegurar y proteger los datos que procesan. Minimizar la cantidad de datos que recopilan es una de las maneras más respetuosas de los derechos humanos para evitar violaciones a la privacidad y otros daños.

# MINIMIZACIÓN DE DATOS EN LA PRÁCTICA: CASOS DE USO COMPLEJOS

Si bien la minimización de datos es un principio claro, los detalles son complicados. A continuación, exploramos la manera en que los principios de minimización de datos deberían aplicarse en el contexto en que se abordan los daños a los derechos civiles, se reduce el impacto negativo de la publicidad conductual y se mejoran los sistemas de *machine learning* (ML).

## Permitir que las organizaciones recopilen datos a fin de identificar violaciones a los derechos civiles y en beneficio de las poblaciones subrepresentadas

Las organizaciones deberían auditar frecuentemente sus sistemas para asegurarse de que están limitando los datos que recopilan solo a los que son necesarios para prestar su servicio, limitando, así, el daño que podrían causar. De hecho, podrían entrar en conflicto con las leyes nacionales si no lo hicieran.[36]

La minimización de datos no impide la recopilación de datos como la raza y el género cuando estos sean necesarios para la prestación del servicio. Sin embargo, se podría argumentar que evita que las organizaciones recopilen información sobre raza, género u otros atributos protegidos (1) para estudiar si la organización quebranta o facilita el quebrantamiento de las leyes de derechos civiles y (2) para beneficiar a poblaciones con insuficiente representación, como las comunidades de color. Sin dicha información, a una organización (y a auditores externos y organismos de control) le resultaría difícil determinar si sus prácticas no protegen los derechos civiles.

[35] Zach Whittaker, *Amazon says government demands for user data spiked by 800% in 2021*, Tech Crunch (1 de febrero de 2021), https://techcrunch.com/2021/02/01/amazon-government-demands-spiked.
[36] La Comisión Federal de Comercio de los Estados Unidos declaró recientemente que el *machine learning* sesgado va en contra de la ley. Elisa Jillson, *Aiming for truth, fairness, and equity in your company's use of AI*, Federal Trade Commission (19 de abril de 2021), https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.

En ese caso, recomendamos elaborar una excepción acotada en torno a los estrictos requisitos de minimización de datos para estos fines.[37] Las organizaciones deberían tener permitido recopilar datos de clases protegidas cuando su propósito sea abordar sus propias prácticas discriminatorias, mitigar o eliminar los daños, o beneficiar a poblaciones subrepresentadas.

**RECOMENDACIÓN**

**En el contexto de un sólido marco de protección de datos personales, debería permitirse que las organizaciones recopilen datos sobre clases protegidas a fin de llevar a cabo auditorías de derechos civiles o de beneficiar a las poblaciones subrepresentadas.**

Una legislación de privacidad completa debería incluir requisitos estrictos de minimización de datos, con una excepción específica para "el propósito de publicitar, comercializar o captar oportunidades económicas para las poblaciones subrepresentadas de una manera justa, que no sea engañosa ni abusiva" y para "legitimar pruebas internas para prevenir la discriminación ilícita o para determinar el alcance o efectividad del cumplimiento [de la organización]" de las leyes de derechos civiles.[38] Con dichas excepciones, las organizaciones podrían auditar sus sistemas para identificar si existen sesgos, pero seguirían estando sujetas a los requisitos de minimización de datos.

Por descontado, una vez que los datos de las clases protegidas sean recopilados y almacenados, no deberían usarse de ninguna otra forma, y deberían estar protegidos de manera estricta contra el acceso no autorizado, la divulgación no autorizada, y otras violaciones a la protección de datos personales. Existen demasiados ejemplos de daños a los derechos civiles provocados por tecnología automatizada, como el hecho de que Facebook permitiera a los anunciantes dirigirse a públicos en función de categorías protegidas,[39] y los públicos similares de Facebook que recrean el sesgo en los conjuntos de datos que brindan sus anunciantes.[40] Ninguna organización debería crear sistemas que discriminen en función de clases protegidas.

## Reducir los daños de la publicidad conductual al limitar la recopilación de datos con fines publicitarios

La publicidad conductual predomina como modelo de negocio en línea. Por lo general, se define como publicidad dirigida a personas en función de sus conductas pasadas. Esta focalización implica el

[37] Todo intento de abordar o "reparar" sistemas que se hayan identificado como sesgados debería regirse por un principio estricto de minimización de datos y no debería permitir la excesiva recopilación de datos de individuos protegidos.
[38] *Consulte The Online Civil Rights and Privacy Act of 2019*, Free Press y Lawyers' Committee for Civil Rights Under Law, https://www.freepress.net/sites/default/files/2019-03/online_civil_rights_and_privacy_act_of_2019.pdf, en la Sección 3(g).
[39] Julia Angwin *et al*., *Facebook (Still) Letting Housing Advertisers Exclude Users by Race*, ProPublica (21 de noviembre de 2017), https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin.
[40] Muhammad Ali *et al*., *Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes*, https://arxiv.org/pdf/1904.02095.pdf.

rastreo de las actividades en línea de las personas, casi siempre mediante el historial de su navegador web, el uso de aplicaciones y otros atributos.[41] También conlleva hacer un "perfilamiento",[42] que se define en EE. UU. como "toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física", en particular, las preferencias, el comportamiento y la ubicación de la persona.[43]

Son muchas las razones por las que se ha criticado el modelo de negocio de publicidad conductual: es invasivo y algo siniestro, puede llevar al perfilamiento y segmentación de personas, y aumenta los riesgos de discriminación.[44] Con el aumento de los argumentos en contra de la publicidad conductual,[45] el movimiento para prohibirla también crece.[46] Existe al menos un proyecto de ley en EE. UU., la Ley de Rendición de Cuentas y Transparencia en torno a los Datos del 2020, del senador Brown, que lograría precisamente eso.[47] En países en los que la minimización de datos ya es un requerimiento, como en la UE gracias al Reglamento General de Protección de Datos, la publicidad conductual ya puede estar restringida significativamente debido a las fuertes disposiciones de suscripción voluntaria en la ley.[48]

Mientras que la evidencia sobre los daños abunda, casi no hay pruebas de los beneficios de que las empresas empleen la publicidad conductual. Puede que no sea tan efectiva como afirman. Un estudio reciente descubrió que los anunciantes retienen solo el 4 % del incremento de las ganancias a partir de la publicidad conductual.[49] En el 2019, cuando *The New York Times* interrumpió los intercambios publicitarios y recurrió a la publicidad contextual, percibió un aumento en sus ganancias.[50] Datos de la empresa de radiodifusión holandesa NPO señalan que, tras dejar la publicidad conductual y reemplazarla por anuncios contextuales en todos sus sitios durante la primera mitad del 2020, sus

[41] El estado de California la definió como "dirigir publicidad a una persona en función de su información personal obtenida de su actividad de consumo en todos los negocios, mayoritariamente en determinados sitios web, aplicaciones o servicios, en lugar del negocio, sitio web específico, aplicación o servicio con los que la persona interactúa de manera intencional". CPRA, sección 1798.140(k), https://thecpra.org.
[42] Francesco Banterle, *Early thoughts on behavioral advertising and the GDPR: a matter of discrimination?*, IP Lens (19 de septiembre de 2017), https://iplens.org/2017/09/19/early-thoughts-on-behavioral-advertising-and-the-gdpr-a-matter-of-discrimination.
[43] Artículo 4.4 del RGPD, https://gdpr-info.eu/art-4-gdpr.
[44] *Chao & Null,* Paying for Privacy, https://www.newamerica.org/oti/reports/paying-our-privacy-what-online-business-models-should-be-limits.
[45] Natasha Lomas, *The case against behavioral advertising is stacking up,* Tech Crunch (20 de enero de 2019), https://techcrunch.com/2019/01/20/dont-be-creepy; Gilad Edelman, *Why Don't We Just Ban Targeted Advertising?* Wired Magazine (22 de marzo de 2020), https://www.wired.com/story/why-dont-we-just-ban-targeted-advertising.
[46] Ban Surveillance Advertising, https://www.bansurveillanceadvertising.com.
[47] Data Accountability and Transparency Act of 2020, https://www.banking.senate.gov/imo/media/doc/Brown%20-%20DATA%202020%20Discussion%20Draft.pdf.
[48] Francesco Banterle, *Early thoughts on behavioral advertising and the GDPR: a matter of discrimination?*, IP Lens (19 de septiembre de 2017), https://iplens.org/2017/09/19/early-thoughts-on-behavioral-advertising-and-the-gdpr-a-matter-of-discrimination.
[49] Natasha Lomas*, Targeted ads offer little extra value for online publishers, study suggests*, Tech Crunch (31 de mayo de 2019), https://techcrunch.com/2019/05/31/targeted-ads-offer-little-extra-value-for-online-publishers-study-suggests.
[50] Jessica Davies*, After GDPR, The New York Times cut off ad exchanges in Europe — and kept growing ad revenue*, Digiday (16 de enero de 2019), https://digiday.com/media/gumgumtest-new-york-times-gdpr-cut-off-ad-exchanges-europe-ad-revenue. Si bien *The New York Times* no es el único anunciante, sí abre un camino posible hacia la publicidad contextual.

ganancias aumentaron mes a mes.[51] Además, la recopilación, la retención y el almacenamiento de datos reunidos para la publicidad también acarrean costos.[52] En lugar de invertir en la elaboración de alternativas que protejan la privacidad, "[m]ucha de la innovación [en la publicidad] se ha abocado específicamente a mostrar anuncios a personas en función de su comportamiento en lugar del contexto del sitio web en el que se encuentran".[53]

De más está decir que la publicidad conductual representa un problema para la minimización de datos. ¿Cómo puede una organización minimizar los datos que recopila si usa dichos datos para alimentar su sistema de publicidad conductual? Si la publicidad conductual es la fuente principal de ingresos de una organización, esta podría pensar que "necesita" absolutamente todos los datos de sus usuarios y usuarias para asegurarse de mostrar los anuncios más relevantes. Es notable que estas organizaciones rara vez exploren si un procesamiento de datos menos intrusivo podría satisfacer sus objetivos, aunque deberían hacerlo para cumplir con los principios de necesidad y proporcionalidad en la UE.[54] Si se vieran confrontados sobre la violación de los principios de minimización de datos, los desarrolladores de la aplicación "Brightest Flashlight" podrían decir que recopilaban datos de geolocalización para mostrar anuncios en función de la ubicación como fuente de ingresos, y eso, hipotéticamente, podría satisfacer un débil o inexistente requisito de minimización de datos, lo que pasaría por alto, además, los problemas de transparencia. Facebook permite que los anunciantes de productos o servicios se dirijan a cualquiera de sus usuarios/usuarias o grupos, lo que implica que la empresa puede recopilar, y generalmente lo hace, cantidades aparentemente ilimitadas de datos de dichas personas.

De no prohibir la publicidad conductual —aunque la prohibición traería grandes beneficios de privacidad—, creemos que, al menos, los entes reguladores deberían intervenir para garantizar que las organizaciones reduzcan los daños provocados por la publicidad conductual. La minimización de datos y los límites de retención cumplen una función importante en la reducción de daños.

**RECOMENDACIÓN**

**Aquellos entes reguladores que no prohíban la publicidad conductual deberían, como mínimo, imponer límites a la recopilación de datos para este propósito.**

[51] Natasha Lomas, *Data from Dutch public broadcaster shows the value of ditching creepy ads*, Tech Crunch (24 de julio de 2020), https://techcrunch.com/2020/07/24/data-from-dutch-public-broadcaster-shows-the-value-of-ditching-creepy-ads/?guccounter=1.
[52] Christopher Tozzi, *5 hidden costs of big data*, Precisely (8 de junio de 2020), https://www.precisely.com/blog/big-data/the-hidden-costs-of-big-data.
[53] *Chao & Null,* Paying for Privacy en la página 12, https://www.newamerica.org/oti/reports/paying-our-privacy-what-online-business-models-should-be-limits/legislation-could-promote-privacy-protective-business-models (cita de Megan Gray, asesora jurídica de DuckDuckGo). *Consulte, además,* Gabriel Weinberg, *What if we all just sold non-creepy advertising?*, N.Y. Times (19 de junio de 2019), https://www.nytimes.com/2019/06/19/opinion/facebook-google-privacy.html.
[54] Necessity & Proportionality, European Data Protection Supervisor, https://edps.europa.eu/data-protection/our-work/subjects/necessity-proportionality_en.

En particular, una organización que recopile datos para propósitos publicitarios, lo cual debería estar minimizado en cumplimiento de los principios de minimización de datos, debería estar obligada a eliminar (no solo desidentificar) dicha información, así como también toda la información inferida a partir de esos datos, pasados los 30 días. Los datos viejos sobre los "intereses" de una persona en base a su historial de búsqueda no solo pueden ser incorrectos, sino también obsoletos, y es probable que quienes vean anuncios basados en su antiguo historial de búsqueda los perciban como intrusivos y "siniestros". Es poco probable que las personas interactúen con anuncios que se perciben de esta forma.[55] Estas limitaciones no eliminarán los daños asociados con la retención indefinida de datos y el perfilamiento de usuarios y usuarias, pero es muy probable que los reduzcan.

## Usar una minimización de datos bien pensada para crear mejores sistemas de *machine learning*

Entrenar sistemas de *machine learning* (ML) puede requerir conjuntos de datos extremadamente grandes.[56] La complejidad de algunos enfoques hacia el *machine learning* puede dificultar determinar si operan en consonancia con el principio de minimización de datos.[57]

Los sistemas de ML son mejores cuando se entrenan usando *buenos* datos, en lugar de simplemente la mayor cantidad de datos posible.[58] Existe una presunción errónea de que siempre es mejor tener tantos datos como sea posible, y de que todos los datos resultan útiles. Sin embargo, quienes crean los sistemas de ML deberían tener en cuenta los principios de minimización de datos, ya que los datos que no son adecuados pueden afectar el rendimiento de un sistema y traer problemas. Se pueden crear técnicas que preserven la privacidad, como la minimización, e incorporarlas en varias fases del *machine learning*.[59] Estas técnicas permiten que, en la elaboración del ML, las decisiones sean intencionales, meticulosas y selectivas en torno a los datos que se recopilan, garantizando que se minimicen los riesgos a la privacidad y se cree, a la vez, un sistema eficaz y funcional.

Por lo tanto, quienes lo desarrollen, deberían intentar garantizar que los datos recopilados ayuden al sistema de ML a tener un buen rendimiento y esto se lleve a cabo en cumplimiento de los derechos humanos y de manera ética.[60] Uno de los logros más avanzados del ML hasta ahora es el sistema de

---

[55] Leslie K. John *et al*., *Ads That Don't Overstep: How to make sure you don't take personalization too far*, Harvard Business Review Magazine (enero-febrero de 2018), https://hbr.org/2018/01/ads-that-dont-overstep (donde se indica que las personas tienen menos probabilidades de interactuar con publicidades que utilizaron información inferida o información recopilada por terceros).
[56] Daniel Leufer *et al*., *AI Myths - AI can solve any problem,* https://www.aimyths.org/ai-can-solve-any-problem.
[57] Abigail Goldsteen *et al*., *Data Minimization for GDPR Compliance in Machine Learning Models*, Cornell University (2020), https://arxiv.org/pdf/2008.04113.pdf.
[58] *Consulte, en general,* Eliza Strickland, *OpenAI's GPT-3 Speaks! (Kindly Disregard Toxic Language*), IEEE Spectrum (1 de febrero de 2021), https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/open-ais-powerful-text-generating-tool-is-ready-for-business.
[59] *Better Machine Learning Through Data Minimization,* Privatar (5 de marzo de 2020), https://www.privitar.com/blog/better-machine-learning-through-data-minimization.
[60] Bernard Marr, *Why AI Would Be Nothing Without Big Data*, Forbes (9 de junio de 2017), https://www.forbes.com/sites/bernardmarr/2017/06/09/why-ai-would-be-nothing-without-big-data.

generación de lenguaje de OpenAI, GPT3, que produjo sorprendentes ejemplos de lenguaje "que suena natural". Sin embargo, en parte porque recibió entrenamiento de una enorme base de datos de texto extraído desde los más recónditos rincones de internet, incluido Reddit, ha reproducido, de manera constante, lenguaje racista, machista y problemático en otros aspectos. Si bien el "big data" dio paso a un desempeño admirable en algunas tareas, también ha llenado al modelo de prejuicios muy preocupantes.[61]

**RECOMENDACIÓN**

**Los desarrolladores de sistemas de *machine learning* deberían adoptar un método para llevar a cabo una minimización de datos para modelos de ML que minimice los efectos en el rendimiento del modelo y salvaguarde los derechos de privacidad.**

La minimización de datos no implica que los desarrolladores no puedan recopilar ciertos tipos de datos. Más bien, requiere que las organizaciones recopilen solo lo que sea necesario para brindar su servicio o producto. Por ejemplo, una herramienta de ML diseñada para aumentar la diversidad racial y de género en un proceso de contratación podría tener la necesidad de recopilar información demográfica sobre la raza y el género. Los desarrolladores deberían también tener permitido recopilar este tipo de información para identificar o auditar áreas de sesgo en los sistemas de ML, como mencionamos anteriormente.

[61] Strickland, *OpenAI's GPT-3 Speaks!*.

# CONCLUSIÓN Y RECOMENDACIONES

La minimización de datos es fundamental para la protección de la privacidad y la reducción de daños en torno a la privacidad. Los datos que no se recopilan no pueden ser utilizados para dañar a las personas. Sin minimizar los datos que se recopilan, los daños a la privacidad continuarán agravándose.

Las siguientes son nuestras recomendaciones:

- ➔ **En el contexto de un sólido marco de protección de datos personales, debería permitirse que las organizaciones recopilen datos sobre clases protegidas a fin de llevar a cabo auditorías de derechos civiles o de beneficiar a las poblaciones subrepresentadas**: las organizaciones deberían tener permitido recopilar datos de clases protegidas cuando su propósito sea abordar sus propias prácticas discriminatorias, mitigar o eliminar los daños, o beneficiar a poblaciones subrepresentadas.

- ➔ **Aquellos reguladores que no prohíban la publicidad conductual deberían, como mínimo, imponer límites a la recopilación de datos para este propósito:** en particular, una organización que recopile datos para propósitos publicitarios, lo cual debería estar minimizado en cumplimiento de los principios de minimización de datos, debería estar obligada a eliminar (no solo desidentificar) dicha información, así como también toda la información inferida a partir de esos datos, pasados los 30 días.

- ➔ **Los desarrolladores de sistemas de *machine learning* deberían adoptar un método para llevar a cabo una minimización de datos para modelos de ML que minimice los efectos en el rendimiento del modelo y salvaguarde los derechos de privacidad**: quienes lo desarrollen, deberían intentar garantizar que los datos recopilados ayuden al sistema de ML a tener un buen rendimiento y esto se lleve a cabo en cumplimiento de los derechos humanos y de manera ética.

**Para obtener más información, comuníquese con:**

**Equipo de protección de datos personales de Access Now | dataprotection@accessnow.org**
**Contacto de prensa | press@accessnow.org**