RDF

DERECHO DE FAMILIA

REVISTA INTERDISCIPLINARIA
DE DOCTRINA Y JURISPRUDENCIA

INCLUYE VERSIÓN DIGITAL

ABELEDOPERROT

ISSN: 1851-1201
RNPI: 5074814

Dirección, administración y redacción
Tucumán 1471 (C1050AAC)
laley.redaccionjuridica@tr.com

Ventas
CASA CENTRAL
Tucumán 1471 (C1050AAC)
Tel.: 4378-4700 / 0810-266-4444

LOCAL I FACULTAD DE DERECHO - UBA
Figueroa Alcorta 2263 (C1425CKB)
Tel. / Fax: 4806-5106

Atención al cliente: 0810-266-4444
Buenos Aires - Argentina

Hecho el depósito que establece la ley 11.723.
Impreso en la Argentina. Printed in Argentina.

Nota de la Dirección: Las opiniones vertidas en los comentarios firmados son privativas de quienes las emiten.

Se terminó de imprimir en la 1ra. quincena de julio de 2021, en los talleres gráficos de La Ley S.A.E. e I., Bernardino Rivadavia 130, Avellaneda - Provincia de Buenos Aires, República Argentina

RDF DERECHO DE FAMILIA

REVISTA INTERDISCIPLINARIA DE DOCTRINA Y JURISPRUDENCIA

Fundada por Cecilia P. Grosman, Celina A. Perrot
y María Bacigalupo de Girard en 1989

Julio 2021 | 100

DOCTRINA

INCLUYE VERSIÓN DIGITAL

ABELEDOPERROT

Sesgos algorítmicos de género con identidad iberoamericana: las técnicas de reconocimiento facial en la mira

Cecilia C. Danesi (*)

Sumario: I. Introducción.— II. ¿Qué es la inteligencia artificial?— III. Los sesgos algorítmicos y la perspectiva de género.— IV. Las técnicas de reconocimiento facial.— V. El marco jurídico y la propuesta de regulación europea.— VI. Reflexión final.

I. Introducción

¿Discriminan las máquinas? ¿Viviremos en el Gran Hermano o en el panóptico del siglo XXI? ¿Existen los sesgos algorítmicos? ¿Tenemos que temerle al desarrollo tecnológico? O peor aún, ¿tenemos que prohibir la inteligencia artificial? En la era de las preguntas sin respuestas, y especialmente sin legislación, estos serán algunos de los interrogantes sobre los que intentaremos echar luz.

(*) Abogada UBA; magíster en Derecho de Daños, Universidad de Girona (España) con beca de Fundación Carolina (tesina en IA y responsabilidad civil); estudió en las Universidades de Salamanca y Paris 2-Panthéon Assas y fue profesora visitante en las Universidades de Oxford, Perugia, Salamanca y diversas universidades argentinas en grado y posgrado; investigadora y docente en la Facultad de Derecho de la UBA en las asignaturas Obligaciones Civiles y Comerciales e Inteligencia Artificial y Derecho; subdirectora del posgrado del mismo nombre; autora del libro "Accidentes de tránsito" (Ed. Hammurabi) y de varios artículos doctrinarios; coordinadora en IALAB y ENTED; directora de la revista *Inteligencia Artificial, Tecnologías Emergentes y Derecho* (Ed. Hammurabi); subdirectora del Instituto de Derecho de la Salud del Colegio de Abogados de San Isidro; ganadora del Premio de Derecho Privado Castán Tobeñas de la Academia Aragonesa de Jurisprudencia y Legislación; doctoranda del Doctorado Internacional en Derecho del Consumidor de las Universidades de Perugia y Salamanca; web: *www.ceciliadanesi.com*, Instagram: @ceciliadanesi, Linkedin: Cecilia Celeste Danesi.

II. ¿Qué es la inteligencia artificial?

Si bien hoy es uno de los temas "de moda" y muchos habrán escuchado hablar sobre ella, debemos comenzar por esbozar alguna definición. En el pasado, hemos aclarado —reiteradas oportunidades— que no hay consenso en torno del concepto de IA; el paso del tiempo no hizo cambiar las cosas. Les propongo para comenzar con una definición "en criollo" y luego, ir a alguna más técnica.

Como su nombre lo indica, la IA viene a intentar emular/imitar la mente humana, lo cual, a primera vista parece bastante pretensioso: y lo es. Esta tecnología procura simular el funcionamiento de nuestro cerebro de la siguiente manera: tiene un conjunto de datos (si son muchos, hablamos de *big data*), sobre la base de los cuales, un algoritmo hace una predicción. Esas predicciones, además, pueden fundarse también en la recopilación de datos proveniente de la interacción del sistema con el entorno, por ejemplo, mediante cámaras, comentarios de usuarios, etcétera.

Se preguntarán ahora qué es un algoritmo. Un algoritmo es una secuencia de pasos descripta en un lenguaje particular (v.gr., una receta de cocina). Para el caso de un programa o una *app*, el algoritmo se escribirá en un lenguaje de programación, un lenguaje que la computadora en-

tiende y le indica qué tiene que hacer. Pero, al enfocarnos en sistemas de *software* que además de ser *software*, están basados en IA, lo cual "significa que tiene componentes que implementan modelos de IA y otros que son programas "normales". Un modelo de IA se implementa por medio de uno o varios programas, pero no es un simple algoritmo, es un meta-algoritmo. La diferencia clave es que un algoritmo define exactamente el proceso a través del cual se toma una decisión, es decir, el algoritmo codifica la solución a un determinado problema. Mientras que un modelo de IA aprende o infiere cómo tomar dicha decisión, es decir, la secuencia de instrucciones del modelo de IA le indica los pasos que tiene que seguir para encontrar una solución a un determinado problema que se le da como entrada" **(1)**.

La Unión Europea ha emitido diversos documentos relativos a la inteligencia artificial en muchos de los cuales abordó la cuestión atinente a la definición. En la comunicación "Inteligencia artificial para Europa" se sostuvo que aquella se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción —con cierto grado de autonomía— con el fin de alcanzar objetivos específicos **(2)**. Posteriormente, esta fue ampliada y actualizada por el Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, en el siguiente sentido: "Los sistemas de inteligencia artificial (IA) son sistemas de *software* (y en algunos casos también de *hardware*) diseñados por seres humanos que, dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno a través de la obtención de datos, la interpretación de los datos estructurados o no estructurados que recopilan, el razonamiento sobre el conocimiento o el procesamiento de la información derivados de esos datos, y decidiendo la acción o acciones óptimas que deben llevar a cabo para lograr el objetivo establecido. Los sistemas de IA pueden utilizar normas simbólicas o aprender un modelo numérico; también pueden adaptar su conducta mediante el análisis del modo en que el entorno se ve afectado por sus acciones anteriores" **(3)**.

De estas definiciones podemos extraer las siguientes consideraciones para terminar de comprender el funcionamiento de la IA. En primer lugar, el alma de la IA son los datos, es por ello que habremos escuchado que son el petróleo de nuestra era. Los datos estarán presentes no solo en el *data set* inicial con el que se entrene el sistema, sino también a raíz de los datos que va percibiendo con sus interacciones. Otra cuestión relevante en la materia es la supuesta autonomía —lo cual está sumamente discutido— que significa que la IA toma decisiones —en principio— sin injerencia externa. Y por último, la capacidad de autoaprendizaje, esto significa que el sistema crea las reglas con base en los datos que le proporcionamos y luego aplica esas reglas para hacer las predicciones. Además, también aprende de las interacciones que hace una vez que es puesto en circulación, de ahí que en la definición se diga que "también pueden adaptar su conducta mediante el análisis del modo en que el entorno se ve afectado por sus acciones anteriores".

Veamos esto con algunos ejemplos. Hoy sabemos que tenemos IA hasta en la sopa: la IA como gestora del contenido en las redes sociales **(4)**, para predecir nuestras preferencias o si tenemos una enfermedad **(5)**, para decidir si somos candidatos para ingresar en una universidad u

(1) MARTÍNEZ, V. – RODRÍGUEZ, O., "Deconstruyendo la inteligencia artificial", *Inteligencia Artificial, Tecnologías Emergentes y Derecho*, nro. 2, Ed. Hammurabi (en edición).

(2) Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones, denominado "Inteligencia artificial para Europa", del 25/04/2018.

(3) "Una definición de la inteligencia artificial: principales capacidades y disciplinas científicas", Grupo independiente de expertos de alto nivel sobre inteligencia artificial, Unión Europea, junio de 2018.

(4) Sobre este tema ver WIERZBA, S. – DANESI, C., "Violencia en las redes sociales ¿Acciones judiciales o normas y algoritmos como clave para la prevención?", LA LEY 2020-A, 641, disponible al 16/04/2021 en *https://www.ceciliadanesi.com/lecturas*.

(5) "El nuevo sistema con inteligencia artificial que reconoce toses compatibles con COVID-19", disponible al 16/04/2021 en *https://www.buenosaires.gob.ar/laciudad/noticias/la-ciudad-creo-un-sistema-con-inteligencia-artificial-que-reconoce-tosesv*.

obtener un crédito **(6)**, para la agricultura **(7)**, para anticipar problemas de mantenimiento en los puentes **(8)**, y un sinfín de etcéteras. Pero de ese sinfín nos vamos a detener en dos. Por un lado, los vehículos autónomos que utilizan inteligencia artificial para tomar decisiones, como por ejemplo avanzar, retroceder, frenar. La IA permite "leer" y "entender" las imágenes que vemos en la vía pública, v.gr., las señales de tránsito. Por el otro, los sistemas de reconocimiento facial, que será uno de los temas en los cuales nos vamos a detener en esta oportunidad.

III. Los sesgos algorítmicos y la perspectiva de género

Hasta aquí todo color de rosas: un sistema que llamamos "inteligente", capaz de procesar enormes cantidades de datos a una velocidad impensada para el humano y que toma decisiones cuasi de forma autónoma. Pero, como sabemos, no todo lo que brilla es oro...

La definición del término "sesgo" presenta algunas dificultades. En primer lugar, enseñan Tolosa y Dibo que aquel tiene diversos significados y se utiliza con distintos alcances; algunos utilizan sesgo cómo sinónimo de prejuicio y otros como equivalente a estereotipo, o también resulta frecuente hablar de una conducta u opinión sesgada en el sentido de haber sido influenciada o manipulada. En estos casos, la palabra "sesgo" generalmente implica una connotación negativa. Y, en consecuencia, se asume que todo esfuerzo tendiente a eliminar sesgos es deseable e, incluso, eliminarlos completamente es un objetivo a cumplir **(9)**.

Sin embargo, tal como hemos señalado en otra oportunidad **(10)**, el sesgo no tiene por qué considerarse *per se* algo negativo. Allí citamos como ejemplo, una entrevista realizada en el programa radial "Perros de la calle" al doctor Facundo Manes, quien sostiene que los sesgos —en muchos casos— nos permiten sobrevivir. Lo fundamenta con el ejemplo del hombre en la Antigüedad: cuando cazaba en la selva y veía una sombra, no se detenía a pensar qué podía ser, sino que salía corriendo y eso le permitía salvarse la vida. Si nos ponemos a pensar, en nuestro día a día encontramos muchos casos similares como, por ejemplo, algo tan simple como alertarse al escuchar un bocinazo **(11)**.

Por otro lado, entendemos por estereotipo de género a "una opinión o un prejuicio generalizado acerca de atributos o características que hombres y mujeres poseen o deberían poseer o de las funciones sociales que ambos desempeñan o deberían desempeñar. Un estereotipo de género es nocivo cuando limita la capacidad de hombres y mujeres para desarrollar sus facultades personales, realizar una carrera profesional

(6) "Un robot puede decidir si te dan (o no) un crédito", disponible al 16/04/2021 en *https://www.elcorreo.com/economia/tu-economia/robot-pueden-decidir-20190405130150-nt.html*.

(7) "John Deere y la inteligencia artificial en la agricultura", disponible al 16/04/2021 en *https://www.aecoc.es/innovation-hub-noticias/john-deere-y-la-inteligencia-artificial-en-la-agricultura*.

(8) Recuerdan que en el 2018 se cayó el puente de Génova al norte de Italia (sí, aún se caen puentes como si estuviéramos en la prehistoria). Bueno, no solo se reconstruyó y ya está en funcionamiento, sino que también cuenta con robots dotados con inteligencia artificial que son capaces de monitorear la estructura en tiempo real para anticipar problemas de mantenimiento y controlar el estado de los materiales. Ello, a los fines de prevenir lo sucedido en el pasado ("El nuevo puente de Génova vuelve a unir la ciudad dos años después del derrumbe que dejó 43 muertos", disponible al 16/04/2021 en *https://elpais.com/internacional/2020-08-03/el-nuevo-puente-de-genova-vuelve-a-unir-la-ciudad-dos-anos-despues-del-derrumbe-que-dejo-43-muertos.html*).

(9) TOLOSA, P. – DIBO, C., "Inteligencia artificial, discriminación por género y derecho: viejos problemas, nuevos desafíos", *Inteligencia Artificial, Tecnologías Emergentes y Derecho*, nro. 2, Ed. Hammurabi, Buenos Aires, en edición.

(10) DANESI, C., "Inteligencia artificial y derecho", *Inteligencia Artificial, Tecnologías Emergentes y Derecho*, nro. 1, Ed. Hammurabi, Buenos Aires, 2020.

(11) En el artículo referido en la cita anterior, citamos otros dos ejemplos. Por un lado, el que relata Daniel KAHNEMAN en su libro "Pensar rápido, pensar despacio" en el cual cuenta, según el psicólogo Gary Klein, el caso de un grupo de bomberos que entró a apagar un incendio y los sesgos cognoscitivos de uno de ellos les permitió salvarse. Por el otro, el de la película "Sully" sobre el vuelo 1549 de US Airways y su piloto Chesley "Sully" Sullenberger quien, basado en su vasta experiencia, decidió realizar un aterrizaje de emergencia en el río Hudson debido a que en el despegue una bandada ocasionó daños en ambos motores; ello en contra de lo que indicaba la torre de control.

y tomar decisiones acerca de sus vidas y sus proyectos vitales" **(12)**.

Ahora bien, ¿qué son los sesgos algorítmicos? Como una primera aproximación, se refieren a sistemas de inteligencia artificial que arrojan respuestas parciales, sesgadas, prejuiciosas, distorsionadas. El problema acaece cuando estas respuestas afectan considerablemente los derechos humanos y conducen a afianzar e incrementar las brechas existentes.

Una conceptualización interesante, y a nuestro modo de ver acertada la brinda Wikipedia, que sostiene que "el sesgo algorítmico ocurre cuándo un sistema informático *refleja los valores de los humanos* que están implicados en la codificación y recolección de datos usados para entrenar el algoritmo" **(13)**. Tal como sabemos, la tecnología no está sesgada y no es discriminatoria, "aprende" a serlo por la transferencia de sesgos y valores que realizamos los humanos al crearla. Así, la tecnología no es ni más ni menos que el fiel reflejo de los principios de facto que rigen en una sociedad.

Ahora bien, los sesgos algorítmicos pueden tener distintas manifestaciones según el área que afecten, como por ejemplo, de género, raciales, étnicos, etc. A su vez, pueden tener diversa procedencia; esto, más allá que haya quedado bien en claro que se trata de un traspaso de valores de la persona a la máquina. Es por ello que hablamos de "inyección de sesgos" y pueden producirse de tres maneras: los programadores, los datos de entrenamiento y el aprendizaje, lo que llamaremos "PEA".

En el primer caso (los programadores), nos referimos a aquellas personas que se encargan de todo el ciclo de vida de la inteligencia artificial: el diseño, la creación, el mantenimiento, la actualización, las decisiones, etc. Acá el problema suele ser el siguiente: la falta de diversidad en los equipos que desarrollan tecnología. Sucede que la participación femenina es escasa o, en algunos casos, nula, lo que conduce a que los estereotipos de género no solo estén presentes en el diseño, sino también que jamás puedan ser detectados. El informe de la ONU, "Tackling Social Norms: A game changer for gender inequalities" (marzo 2020) señala que, a pesar de décadas de progreso en cerrar la brecha de igualdad de género, cerca de 9 de cada 10 hombres y mujeres en todo el mundo tienen algún tipo de prejuicio contra las mujeres (a iguales resultados arriba la UNESCO en un informe de enero de 2019). Este tipo de prejuicios son los que se trasladan a los sistemas de IA y, al tener equipos de desarrolladores carentes de diversidad, el sesgo se perpetua.

Romei y Ruggieri proponen cuatro estrategias para prevenir la discriminación en el análisis: 1) la distorsión controlada de los datos de entrenamiento, 2) la integración de criterios de antidiscriminación en el algoritmo de clasificación, 3) los modelos de clasificación post-procesamiento y, 4) la modificación de las predicciones y de las decisiones para mantener la equidad. El Grupo de investigación de Inteligencia Artificial, Filosofía y Tecnología (GIFT) propone que podría agregarse una nueva opción a esa lista como estrategia para prevenir la discriminación: la extrapolación del experimento mental del "velo de la ignorancia" a la hora de diseñar IA. Así, sostienen que "sería enriquecedor que al momento de desarrollar un cierto algoritmo, los diseñadores lo hicieran pensando en desconocer cuál es su identidad en términos de sexo, género, edad, educación, ingresos, talentos o color de piel. Al no saber cómo son ni en qué condiciones vivirán en el mundo, podría favorecerse (al menos en principio) una distribución más justa y equitativa de las oportunidades sin el abierto impacto de sesgos personales" **(14)**.

En el segundo supuesto encontramos los datos de entrenamiento que son aquellos datos de los cuales "aprende" el sistema, es decir, de los cuales extrae patrones. Al igual que sucede con una receta de cocina, si utilizamos ingredientes vencidos o en mal estado, la torta saldrá horrible. Veamos algunos ejemplos. Amazon desa-

(12) ONU, "Los estereotipos de género y su utilización", Oficina del Alto Comisionado para los Derechos Humanos, disponible al 07/05/2021 en *https://www.ohchr.org/sp/issues/women/wrgs/pages/genderstereotypes.aspxv*.

(13) El destacado nos pertenece.

(14) PEDACE, K. – BALMACEDA, T. – PÉREZ, D. – LAWLER, D. – ZELLER ECHENIQUE, M., "Caja de herramientas humanísticas", CETyS, fAIr Lac (BID), disponible al 03/05/2021 en *https://grupo.gift*.

rrolló un sistema de IA para calificar los CV de aspirantes a puestos de trabajo y, finalmente, se comprobó que puntuaba de forma inferior a las mujeres que a los hombres frente a iguales antecedentes profesionales. Esto sucedió porque el algoritmo había sido entrenado con los datos últimos de los últimos diez años de la compañía donde, quienes habían ocupado puestos de liderazgo y, por consiguiente, obtenido buenas calificaciones, habían sido los varones. Muchos estarán pensando una fácil solución: "¡Cecilia, eliminemos el género en el CV y listo!" Pues no es tan sencillo. Resulta que, en el caso de Amazon, los CV no tenían el género, pero el sistema detectaba ciertos patrones que vinculaba con lo femenino (p. ej., pertenecer a una asociación de derechos de las mujeres) y bajaba el puntaje. Pero además de eso, en algunos casos, el género es de vital importancia como sucede en cuestiones médicas. Al respecto, se señala que la neutralidad de género no necesariamente implica trato equitativo dado que, en ciertas situaciones, pueden existir razones plausibles que justifiquen el tratamiento desigual y no se configure un caso de discriminación. Así, se afirma que "una crítica que suele hacerse a ciertos algoritmos que intentan predecir el peligro de fuga para decidir la prisión preventiva, es que usualmente no captan adecuadamente que, en el caso de las mujeres, la tasa de reincidencia es mucho más baja que la de los hombres y, por lo tanto, se reclama que el diseño de los algoritmos debería prever esta distinción. Es decir, se demanda un trato desigual para lograr una solución equitativa **(15)**. Cabe destacar que si bien la elección de los datos está en manos de las personas (programadores) y su diligente depuración puede evitar el sesgo algorítmico, lo separamos del modo de "inyección" anterior para que quede bien en claro de dónde proviene el problema y, en consecuencia, una posible solución.

Con relación a la inyección de sesgos a través de los datos, el informe citado de la ONU refiere que los algoritmos de aprendizaje automático no están sesgados desde su nacimiento, sino que aprenden a ser parciales. El sesgo algorítmico se produce cuando el algoritmo de aprendizaje se entrena en conjuntos de datos sesgados y, posteriormente, aprende "con precisión" los patrones de sesgo en los datos. En algunos casos, las representaciones aprendidas dentro de los algoritmos de aprendizaje automático pueden incluso exagerar estos sesgos.

La tercera área donde se inyectan los sesgos en la IA es en el "aprendizaje", la cual se produce por las interacciones que hace el sistema del entorno una vez puesto en circulación. En este caso, puede que el sistema no haya estado sesgado en su nacimiento, pero al comenzar a funcionar, incorporó los sesgos provenientes de los nuevos *inputs*. Un ejemplo de esto es Tay, el *chatbot* de Microsoft, que fue creado para mantener conversaciones divertidas en redes sociales y a las pocas horas de estar en funcionamiento hacía comentarios racistas y xenófobos, por lo que en menos de 24 horas tuvo que ser deshabilitado.

Cabe aclarar que muchas veces la predicción sesgada proviene de las tres causas mencionadas o también, puede resultar sumamente difícil identificar cuál de ellas fue la responsable. Al respecto, la MIT Technology Review sostiene que normalmente nos limitamos a responsabilizar a los datos de entrenamiento como únicos culpables del sesgo algorítmico; pero la realidad evidencia matices, el sesgo puede aparecer mucho antes que los datos se recopilen y también en muchas otras etapas del proceso de aprendizaje profundo.

En razón de ello, identifican tres etapas claves. La primera de ellas es la "definición del problema", esto hace referencia a que en el momento que los informáticos crean un modelo de aprendizaje profundo, deben decidir cuál es su objetivo. Si una compañía de tarjetas de crédito **(16)**, por ejemplo, necesitará predecir la solvencia de sus clientes, pero sucede que "solvencia" es un concepto bastante difuso con lo cual habrá que tomar decisiones y el problema es que

(15) KLEINBERG *et al.* (2018), citado en TOLOSA, P. – DIBO, C., ob. cit.

(16) Ya que hablamos de tarjetas de crédito, no podemos dejar de comentarles el caso de la tarjeta de crédito Apple, la cual —según denunció David Heinemeier Hansson— es sexista ya que le dio a su esposa un límite de crédito 20 veces menor que a él, frente al mismo patrimonio (más ejemplos en DANESI, C. "The impact of Artificial Intelligence on Women´s Rights: a legal point of view", in *The Fourth Industrial Revolution and its Impact on Ethics - Solving the Challenges of the Agenda 2030*, Springer, 2021).

"esas decisiones se toman por diversas razones comerciales que no son ni la imparcialidad ni la discriminación. La segunda etapa es la "recogida de los datos" y el sesgo aquí se produce por dos vías: o los datos recopilados no son representativos de la realidad o reflejan prejuicios ya existentes (como sucedió en el caso de Amazon comentado precedentemente). Y, por último, la "preparación de los datos" donde se seleccionan los atributos que deseamos que el algoritmo tenga en cuenta. En nuestro caso de modelar la solvencia crediticia, un "atributo" podría ser la edad del cliente, los ingresos o la cantidad de préstamos pagados. Esto es lo que se conoce como el "arte" del aprendizaje profundo: elegir qué atributos considerar o ignorar puede influir significativamente en la precisión de la predicción de un modelo **(17)**.

En conclusión y más allá de las consideraciones técnicas, hay que dejar bien en claro que la tecnología no es neutra, como tampoco la información que recibimos sobre ella. Un interesante estudio que analizó diversos artículos periodísticos sobre inteligencia artificial resaltó que "el uso de un lenguaje inclusivo, o la incorporación de datos desagregados por sexos, así como la crítica a estereotipos de género no han sido lo común en los textos periodísticos o de opinión de los periódicos analizados (González Fernández 2017)". Entre los ejemplos mencionaron que "mientras, los robots de un restaurante indio son 'camareras' con andar elegante (*El País*, 13/01/2018), Estados Unidos doblega a Japón en el primer combate real entre robots (Muela, 2017), y las llamadas 'robots sexuales' invaden el mercado con estereotipos de género muy acusados y vinculados a la industria pornográfica sin que, en la mayoría de los casos, los periódicos consultados aborden el debate ético y filosófico sobre el concepto de 'sexualidad'" **(18)**.

IV. Las técnicas de reconocimiento facial

Para terminar de comprender el funcionamiento de los sesgos algorítmicos y, principalmente, su gravedad, vamos a detenernos en un ejemplo concreto que es uno de los que más preocupa en la actualidad: las técnicas de reconocimiento facial **(19)**. Estas técnicas, que nos invitan a imaginarnos viviendo en Gran Hermano o en el Panóptico, consisten en el uso de tecnología para recopilar nuestros "datos biométricos", es decir, "datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona, como imágenes faciales o datos dactiloscópicos" **(20)**. En pocas palabras, las cámaras toman una imagen en tiempo real, la compara con millones de imágenes en segundos y, con ello, identifica si una persona —por ejemplo— tiene pedido de captura.

Los sistemas de reconocimiento facial están más presentes de lo que pensamos. Desde el celular (para desbloquearlo o los filtros de Instagram), hasta el control de las calles y el acceso a un edificio. Su utilización es de lo más polémica ya que podría afectar de forma directa derechos humanos como la privacidad, libertad de asociación, de reunión y de expresión. Pero además de todo ello, nuevamente aquí aparece el problema de los sesgos.

---

(17) HAO, K., "Cómo se produce el sesgo algorítmico y por qué es tan difícil detenerlo", MIT Technology Review, 08/02/2019, disponible al 04/05/2021 en *https://www.technologyreview.es/s/10924/como-se-produce-el-sesgo-algoritmico-y-por-que-es-tan-dificil-detenerlo*.

(18) TAJAHUERCE ÁNGEL, I. – FRANCO, Y., "Periódicos digitales españoles e información sobre robótica e inteligencia artificial: una aproximación a imaginarios y realidades desde una perspectiva de género", *Revista de Comunicación de la SEECI*, nro. 48, 15/03/2019-15/07/2019, ps. 173-189. Las autoras concluyen que "la ciencia y la tecnología no son neutras, conllevan una fuerte carga política, social y económica e ideológicamente contribuyen a construir mundos más adecuados para el ser humano o lo contrario. Por otro lado, la escasa formación con perspectiva de género en todos los niveles de responsabilidad de las y los profesionales de la información periodística, contribuye a que no se detecten los estereotipos de género en los productos que se distribuyen y presentan en Congresos, Ferias Internacionales, Seminarios de Investigación y Centros de Investigación".

(19) Acá pueden ver un video explicativo del funcionamiento: "¿Cómo funciona la tecnología de identificación facial?", BBC News *https://www.youtube.com/watch?v=2FbQ6RnaI0I*.

(20) Reglamento (EU) 2016/679 del Parlamento Europeo y del Consejo de 27/04/2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la directiva 95/46/CE (GDPR).

Uno de los casos más conocidos, y que de hecho la llevó a la nueva pantalla grande —Netflix **(21)**— es el de la investigadora del laboratorio de medios del MIT, Joy Buolamwini, quien comienza el recomendadísimo documental *Coded Bias* colocando una máscara de color blanco en su rostro para que pueda ser identificado por la tecnología en análisis. Ella comprobó que esos sistemas tienen menor precisión para identificar rostros de piel oscura y/o de mujeres; lo cual constituye una preocupación que comparten muchas organizaciones internacionales **(22)**.

Esa falla en la identificación de determinados grupos de personas es real y palpable. Uno de los casos fue el de Robert Julian-Borchak Williams, afroamericano de 42 años que fue detenido en la puerta de su casa por la policía de Detroit por un delito que no había cometido. Le tomaron fotos, las huellas digitales, el ADN y lo interrogaron sin darle explicaciones, recién 18 horas después le informaron el motivo del arresto. Finalmente, la Fiscalía desestimó el caso por falta de pruebas **(23)**. Detroit no es el único Estado que utiliza este sistema, en otras ciudades y países del mundo, como Argentina y Reino Unido, se usa con los mismos fines, mientras que en otras como San Francisco o Cambridge está prohibido. Diversas asociaciones e investigadores de prestigiosas universidades alzaron su voz frente a esta nueva tecnología que puede llevarnos a que los Gobiernos puedan tener un control absoluto de cada uno de nuestros movimientos y decisiones **(24)**.

Lo que todos y todas queremos evitar está claro: llegar al modelo chino. Este país aplica el "Sistema de crédito" **(25)** el cual consiste en asignarles puntajes a los ciudadanos sobre la base de sus comportamientos, acciones, si pagan o no impuestos o cualquier cosa que hagan que pueda ser vigilada. Esto incluye que la expresión de tus ideas —y las de tus amigos— ya sea políticas o religiosas, deben estar acordes al Gobierno de lo contrario, te bajan puntos. Lo mismo sucede si ponés la música fuerte, no pagas una multa o cometes un delito. Si tenés un puntaje alto, significa que sos confiable para el Estado y, por ello, podrás escoger a qué colegio puedan ir tus hijos, ocupar cargos públicos, viajar, sacar el pasaporte, etc. ¿Cómo lo hace? A través de la tecnología, entre ellas, las técnicas de reconocimiento facial, las que no solo están en las calles sino también en las gafas de la policía **(26)**.

V. El marco jurídico y la propuesta de regulación europea

A esta altura del partido, muchos se preguntarán por el marco legal que acoge las tecnologías de reconocimiento facial. Vamos a analizar dos: la argentina y la propuesta de regulación de la Unión Europea.

V.1. La regulación del reconocimiento facial en la Ciudad Autónoma de Buenos Aires

En la Ciudad Autónoma de Buenos Aires, se utilizan las técnicas de reconocimiento facial

(21) En Netflix podes ver el documental "Coded Bias" que ejemplifica con suma claridad todas las cuestiones abordadas en este artículo.

(22) ONU, Programa de las Naciones Unidas para el Desarrollo, "Tackling Social Norms: A game changer for gender inequalities", marzo 2020 disponible al 05/05/2021 en *http://hdr.undp.org/sites/default/files/hd_perspectives_gsni.pdf*, y European Union Agency for Fundamental Rights, "Facial recognition technology: fundamental rights considerations in the context of law enforcement", 2020, disponible al 05/05/2021 en *https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf*.

(23) "Detenido injustamente un afroamericano en EE.UU. por un error en el sistema de reconocimiento facial", *El País*, 26/06/2020, disponible al 05/05/2021 en *https://elpais.com/tecnologia/2020-06-26/un-afroamericano-es-detenido-injustamente-por-un-error-en-el-sistema-de-reconocimiento-facial.html*.

(24) Georgetown Law, "America Under Watch. Face Surveillance In The United States", *Center on Privacy & Technology*, 16/05/2019, disponible al 05/05/2021 en *https://www.americaunderwatch.com/#recommendations*, Big Brother Watch, Reino Unido, *https://bigbrotherwatch.org.uk* y, "Reclaim your Face", Unión Europea, *https://reclaimyourface.eu/*.

(25) China Copyright and Media, https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/.

(26) Hong Kong viene protagonizando varias protestas para evitar llegar al modelo chino, sobre este tema ver "En las protestas en Hong Kong, los rostros son armas", *The New York Times*, *https://www.nytimes.com/es/2019/07/31/espanol/reconocimiento-facial-hong-kong.html*.

para detectar prófugos de la justicia. Las cámaras que forman parte del Sistema Público Integral de Video-Vigilancia están ubicadas en diversas calles y estaciones del subte porteño, toman imágenes en tiempo real y las comparan en cuestión de segundos con una base de datos pública del Co.Na.R.C. (Consulta Nacional de Rebeldías y Capturas) donde se registran los pedidos de captura de los sujetos prófugos de la justicia.

El sistema fue incorporado mediante la res. 398/19 del Ministerio de Justicia y Seguridad y, en octubre de 2020 se ha reformado la ley sobre "Sistema integral de seguridad pública de la Ciudad Autónoma de Buenos Aires" (5688) incluyendo dicha técnica. Desde que se emitió la resolución, el camino no ha sido pacífico ya que diversas asociaciones alzaron su voz y lo criticaron severamente **(27)**. Los reclamos no quedaron solo en manifestaciones, sino que también llegaron a Tribunales. El Observatorio de Derecho Informático interpuso una acción de amparo contra el Gobierno de la Ciudad de Buenos Aires a efectos que brinde la información solicitada en su pedido de acceso a la información relacionado con la res. 398/MJYSGC/2019, mediante la cual se aprobó la implementación en el ámbito de la Ciudad del Sistema de Reconocimiento Facial de Prófugos. El juzgado resolvió que debía admitirse la acción de amparo ya que, frente al requerimiento de información del amparista sobre el mencionado sistema, la oportunamente suministrada había sido parcial e insuficiente. Por ello, ordenaron al GCBA a que conteste una serie de preguntas; aquí copiamos algunas que nos han llamado la atención: d. ¿A qué se refiere con historia de los eventos? ¿Qué información se almacena? ¿Dónde es almacenada esta información? ¿Quién tiene acceso a esa información y por cuánto tiempo?; f. ¿Cuántos usuarios con los dos distintos permisos existen? ¿Qué cantidad de usuarios están limitados a la visualización de los datos? ¿Cuántos usuarios existen con total disponibilidad para todas las operaciones? ¿Quién otorga estos permisos? ¿De qué manera y con qué criterio se otorgan esos permisos?; h. ¿Qué criterio se utilizó a efectos de considerar que un 15% de falsos positivos era un porcentaje aceptable?; m. ¿Cuánta es la cantidad mínima de personas necesarias para que se dé un caso de "merodeo"? y; q. ¿Con que fin se recolecta la información acerca de la detección de emoción en el rostro de las personas? ¿Por qué se necesita detectar la emoción del rostro de las personas cuando el sistema sería utilizado exclusivamente para la detección de prófugos?

Antes de comentar la reforma de la ley 5688, debemos señalar que al momento del dictado de la res. 398/2019 ya se encontraban vigentes los "Criterios orientadores e indicadores de mejores prácticas en la aplicación de la ley 25.326" (Ley de Protección de Datos Personales), los cuales resultan de gran importancia para el tema bajo estudio. En el primer criterio se establece el "Derecho de acceso a datos personales recolectados mediante sistemas de video vigilancia" y regula el procedimiento a seguir en dicho caso. Por ejemplo, reglamenta que "el responsable de la base de datos debe proporcionar los datos personales en forma clara, acompañados de una explicación del tiempo en que se registró al titular de los datos, lugar en el que el sistema de video vigilancia lo registrara, finalidad, eventuales cesiones y/o destino de los datos". En este punto es importante señalar que este derecho del titular de los datos debe ir acompaño del derecho de rectificación y el de remoción de sus datos, pues de lo contrario la disposición carece de sentido.

El criterio dos, en cambio, establece un derecho que es común a cualquier sistema de tratamiento automatizado de datos (básicamente, sistemas de IA), que prescribe que cuando ese tratamiento conlleve "efectos jurídicos perniciosos o lo afecten significativamente de forma negativa, el titular de los datos tendrá derecho a solicitar al responsable de la base de datos una explicación sobre la lógica aplicada en aquella decisión". Aquí encontramos una cuestión angular en materia de inteligencia artificial pues en muchos casos —especialmente en *Deep Learning*— los modelos resultan no interpretables, es decir, que no se puede leer y determinar porque se llegó a una predicción y no a otra.

(27) Asociación por los Derechos Civiles (ADC), "#ConMiCaraNo: Reconocimiento facial en la Ciudad de Buenos Aires", 23/05/19, disponible al 06/05/2021 en *https://adc.org.ar/2019/05/23/con-mi-cara-no-reconocimiento-facial-en-la-ciudad-de-buenos-aires* y "Reconocimiento facial: críticas al sistema y advertencias por riesgos para la privacidad", *Diario Perfil*, disponible al 06/05/2021 *https://www.perfil.com/noticias/politica/sistema-reconocimiento-facial-gobierno-ciudad-criticas-y-advertencias-por-riesgos-para-privacidad.phtml*.

Esto puede tener efectos nocivos si pensamos en sistemas de IA que nos deniegan el acceso a una oportunidad o un derecho. Es por ello que, la mayoría de los principios éticos de IA hablan de explicabilidad, transparencia y cajas blancas.

Pasemos al criterio cuatro que es sumamente crítico. Primero define a los datos biométricos como "datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona humana, que permitan o confirmen su identificación única". Y con relación a ello, regula que los que "identifican a una persona se considerarán datos sensibles (conforme el art. 2º, ley 25.326) únicamente cuando puedan revelar datos adicionales cuyo uso pueda resultar potencialmente discriminatorio para su titular (v.g. datos que revelen origen étnico o información referente a la salud)". Si bien el texto no implica que los datos biométricos carezcan de protección, al quitarles la categoría de sensibles, se los inhibe de una tutela mayor.

Por último, el criterio cinco se aboca al consentimiento (también al de niñas, niños y adolescentes) y establece que "en relación a la cesión de datos personales entre organismos públicos, no se requiere el consentimiento del titular de los datos y se cumple con las condiciones de licitud, en la medida en que (i) el cedente haya obtenido los datos en ejercicio de sus funciones, (ii) el cesionario utilice los datos pretendidos para una finalidad que se encuentre dentro del marco de su competencia y, por último, (iii) los datos involucrados sean adecuados y no excedan el límite de lo necesario en relación a esta última finalidad". Tal como podemos apreciar, la norma abre un amplio abanico de posibilidades que facilita que nuestros datos circulen por la administración pública casi sin inconvenientes.

V.2. La propuesta de regulación europea

En abril de 2021 la Unión Europea dio un gran paso para conseguir una inteligencia artificial respetuosa de los derechos, presentó la propuesta *Artificial Intelligence Act*. En el eje central, encontramos la clasificación de la inteligencia artificial por niveles según el riesgo que presenten a los derechos de las personas. Veamos de qué se trata.

En el primer nivel se sitúan los sistemas que están prohibidos, de los cuales vamos a destacar dos sobre la base de nuestro objeto de análisis. Por un lado, prohíbe que las autoridades públicas utilicen sistemas de IA que evalúen o clasifiquen a la confiabilidad de las personas físicas en función de su comportamiento social o personal que conduzca a: 1) un trato desfavorable de esas personas en contextos sociales donde no se han recopilado los datos originariamente y/o, 2) un trato desfavorable que sea injustificado o desproporcionado con respecto a su comportamiento social o su gravedad. En pocas palabras, la Unión Europea veda la implementación del sistema de crédito social chino que vimos en el pto. iv).

Por otro lado, también se prohíbe el uso de sistemas de identificación biométrica remota "en tiempo real" (ergo, los sistemas de reconocimiento facial) en espacios de acceso público con el propósito de hacer cumplir la ley, excepto que dicho uso sea estrictamente necesario para uno de los siguientes objetivos: i) la búsqueda selectiva de posibles víctimas específicas de delitos, incluidos los niños desaparecidos; ii) la prevención de una amenaza específica, sustancial e inminente a la vida o la seguridad física de personas físicas o de un ataque terrorista; iii) la detección de un sospechoso o autor de un delito incluido en el art. 2º de la decisión 2002/584/JHA62 **(28)**. La crítica que se le ha hecho a esta norma es que peca de permisiva pues, si bien como regla general prohíbe los sistemas de reconocimiento facial en tiempo real en espacios públicos, prevé un listado amplio de excepciones.

En el art. 6º se contemplan los sistemas considerados de alto riesgo, los que sí podrán ser utilizados pero siempre y cuando cumplan con determinados requisitos: a) establecer, implementar y documentar un sistema de gestión de riesgos; b) datos y gobernanza de datos de los datos de entrenamiento, validación y prueba, esto incluye una evaluación previa de la disponibilidad, cantidad e idoneidad de los conjuntos de datos que se necesitan y evaluación de los posibles sesgos, elemento esencial para evitar desigualdades y discriminación (ya hemos visto precedentemente el rol fundamental

(28) En el párrafo siguiente establece los requisitos que se deben cumplir para poder utilizar los sistemas de IA en los supuestos de las excepciones.

que tienen los datos en materia de sesgos algorítmicos); c) capacidad de almacenar y registrar los eventos mientras están en funcionamiento (lo que luego servirá como prueba) y garantizar la trazabilidad; d) garantizar la trasparencia en sus operaciones que permita al usuario interpretar el resultado (*output*) y usarlos apropiadamente; e) la supervisión humana y; f) precisión, robustez y ciberseguridad.

Para determinar si un sistema es de alto riesgo hay dos posibilidades: o que encuadre en el párrafo 1º art. 6º de la propuesta de reglamento (se deben cumplir las dos condiciones) o bien estar comprendidos en el Anexo III **(29)**.

Finalmente, están los sistemas de riesgo limitado, como los agentes conversacionales o las *deep fakes* (en estos casos se exigen requisitos mínimos como informar a la persona que está interactuando con un sistema de IA) y; los de riesgo mínimo que deben respetar la legislación vigente sin normas adicionales.

VI. Reflexión final

Tal como hemos visto a lo largo de este artículo, la tecnología posee un impacto directo en casi todos los ámbitos en los que se desarrolla una persona, pero el problema aparece cuando aquella le impide acceder a un determinado derecho o bien su utilización conlleva a la violación de un derecho humano. En el caso de las técnicas de reconocimiento facial utilizado por los Estados para hacer cumplir la ley (p. ej., para la detección de prófugos de la justicia), la posible afectación es sumamente peligrosa ya que puede englobar discriminación, y afectación a la privacidad, a la intimidad, a la libertad de asociación, de expresión, etcétera.

La cuestión toma un tinte aún más oscuro cuando la vulneración atenta contra grupos vulnerables pues su marginalidad es tal que ni siquiera hay datos representativos y fiables de esos grupos y además, los datos sesgados vuelven a entrar al sistema reforzando el sesgo; lo que se conoce como "bucle de retroalimentación pernicioso" **(30)**. Así, los sesgos algorítmicos tienen la capacidad de afianzar y ensanchar las brechas de desigualdad existentes. Entre esos colectivos, encontramos a las mujeres con lo cual, la perspectiva de género debe incorporarse en el desarrollo de la tecnología. Y no simplemente al momento del diseño, comercialización, puesta en funcionamiento, etc., sino también desde la educación; desde que formamos a quienes en el futuro trabajarán con la tecnología.

Y mientras tanto, nos preguntamos, ¿qué sucede en el seno de las empresas? ¿Están preocupadas por los sesgos algorítmicos? Como para muestra basta un botón, vamos con el caso de Google que en menos de tres meses decidió despedir a Timnit Gebru y a Margaret Mitchell, ambas investigadoras del departamento de IA ética de Google. Gebru es co-creadora de "Black in AI", una comunidad de investigadores negros que trabajan en IA. Fue despedida porque se negó a retractarse de un artículo académico de su autoría sobre el enorme desperdicio de energía de entrenar modelos de IA y el sesgo existente dentro de los sistemas que se entrenan con palabras que se encuentran en Internet. Mitchell, por su parte, formaba parte de la extensa lista de personas que se quejaron del despido de Gebru. Ella también ha sido despedida.

En suma, reiteramos una vez más que el *quid* de la cuestión para eliminar sesgos no está en prohibir la innovación, sino está en nosotros, los humanos, que impregnamos la tecnología con nuestros prejuicios. Sociedades justas crean, avanzan y se desarrollan en marcos justos y equitativos.

(29) El anexo III se enumeran los sistemas de alto riesgo según el área donde se utilizan, a saber: identificación biométrica y categorización de personas humanas; gestión y operación de infraestructura crítica; educación y formación profesional; empleo, gestión de trabajadores y acceso al autoempleo; acceso y disfrute de los servicios públicos y privados esenciales; cumplimiento de la ley; gestión de migración, asilo y control de fronteras; administración de justicia y procesos democráticos.

(30) O'NEIL, C., "Armas de destrucción matemáticas", *Capitan Swing*, 2017.