GuIA.ai

Videovigilancia con reconocimiento facial, inteligencia artificial y derechos humanos: ni apocalipsis ni utopía

José Fernando Flórez Ruiz y César Orlando Díaz Benito

CETyS

LATAM DIGITAL
CENTRO DE POLÍTICA DIGITAL PARA AMÉRICA LATINA

Videovigilancia con reconocimiento facial, inteligencia artificial y derechos humanos: ni apocalipsis ni utopía

Autores: José Fernando Flórez Ruiz y César Orlando Díaz Benito

Diseño: Mónica Castellanos

Videovigilancia con reconocimiento facial, inteligencia artificial y derechos humanos: ni apocalipsis ni utopía

Noviembre 2021

José Fernando Flórez Ruiz

Abogado de la Universidad Externado de Colombia, especialista en Derecho Constitucional, magíster en Políticas Públicas y Ph. D. en Ciencia Política de la Universidad París II Panthéon-Assas. Fulbright Visiting Scholar de Johns Hopkins University. Ha sido profesor, conferencista e investigador visitante en varias universidades e instituciones del mundo. Ha trabajado como asesor y funcionario en el Senado de la República, la Presidencia de la República, la Alcaldía Mayor de Bogotá y la Fiscalía General de la Nación en Colombia. Actualmente es el Director de Planeación de la Registraduría Nacional del Estado Civil de Colombia, donde creó y lidera el Grupo de Analítica de Datos e Inteligencia Artificial de la entidad. Twitter: @florezjose

César Orlando Díaz Benito

Ingeniero eléctrico de la Universidad de los Andes, magíster en Ingeniería Electrónica de la Pontificia Universidad Javeriana, Ph. D. en Informática de la Universidad de Luxemburgo y estudios de posdoctorado de la Universidad de los Andes. Se ha desempeñado como profesor e investigador en diferentes universidades de Colombia en las áreas de redes de computadores, sistemas distribuidos, seguridad informática, programación, aprendizaje de máquina e inteligencia artificial. Actualmente se desempeña como líder del Grupo de Analítica de la Superintendencia de Industria y Comercio, Delegatura para la Protección de la Competencia. También es profesor de planta de la Universidad ECCI, consultor en *data analytics*, algoritmos de inteligencia artificial, *cloud*, HPC y algoritmos *energy-efficient* para *scheduling* y *resource allocation*.

> *"Depending on whom you ask, the increased prevalence and capabilities of CCTV surveillance could make society safer and more efficient, could trample on our rights to privacy and freedom of movement, or both. No matter which side you argue, the fact is that live video surveillance is ramping up worldwide"*
>
> ***- Paul Bischoff***

Resumen ejecutivo

Esta investigación analiza el impacto que la videovigilancia con reconocimiento facial tiene sobre los derechos humanos. Para ello, parte de dos supuestos: por un lado, la urgencia de superar las narrativas extremas, bien sea apocalípticas o irrestrictamente optimistas, acerca del impacto de la inteligencia artificial (IA) sobre los derechos humanos (DDHH); y por el otro, la necesidad de dejar atrás un enfoque limitado a la ética como único marco regulatorio a partir del cual se establecen límites al uso de la IA. Para ello, retoma la periodización propuesta por Webster (2004), para quien las respuestas a los desafíos planteados por la emergencia de la tecnología de videovigilancia han seguido tres grandes etapas: la etapa de la "innovación", centrada en la auto-regulación, las normas no vinculantes y los marcos éticos; la etapa del "despegue" y la etapa de la "sofisticación", caracterizada por la legislación formal sobre la tecnología. Consideramos que el fortalecimiento de las respuestas basadas en el derecho, con su carácter vinculante, puede contribuir a acelerar la curva de aprendizaje de los países en desarrollo, que actualmente siguen trayectorias regulatorias a diversas velocidades. Nuestra indagación responde los siguientes interrogantes: 1) ¿Cómo funciona la "geopolítica de la IA", es decir, dónde se producen las tecnologías de IA y cuáles son sus principales centros de consumo? 2) ¿Qué consecuencias tiene sobre las caracterizaciones hegemónicas y la utilización de la IA que la mayor parte de esta tecnología sea producida por unas pocas empresas localizadas en los países más desarrollados? 3) ¿Qué respuestas se han dado a nivel internacional, desde el derecho y más allá de la ética, frente a los desafíos planteados por la IA? 4) ¿Cuáles son los principales riesgos que la videovigilancia con reconocimiento facial plantea en materia de DDHH? 5) ¿Cuál es el marco normativo idóneo en Colombia para mitigar los principales riesgos de vulneración de los derechos humanos?

Contenido

1. Introdución

Hace una década, en un libro titulado *Nuestro futuro biométrico*, Kelly Gates (2011) anticipaba el advenimiento de una nueva cultura de la supervigilancia, potencializada por el desarrollo de las *tecnologías de reconocimiento facial* (TRF en lo sucesivo), que obligaría a nuestras sociedades a adelantar difíciles discusiones de política pública en términos de compensaciones entre la seguridad y la privacidad de las personas. Para la época, la autora escribió que estaba por verse "si en realidad los sistemas de reconocimiento facial automatizado y de análisis de la expresión pueden lograr todo lo que sus promotores buscan" (p. 5).

Los acontecimientos de los últimos diez años les dan la razón a quienes apostaron por la rápida expansión de las TRF a diversos campos de la vida social, pero sigue abierta la pregunta sobre si el perfeccionamiento de estas tecnologías les permitirá cumplir todas sus promesas de impacto positivo en la vida de las personas. Entre otras razones, porque la expansión de la biometría facial también conlleva altos riesgos de afectación de los derechos humanos que urge mitigar.

Para septiembre de 2019, 64 países del mundo, entre los 176 considerados, utilizaban sistemas de reconocimiento facial para labores de videovigilancia (Feldstein, 2019)[1]. Para 2022, se proyecta que en China habrá una cámara de circuito cerrado de televisión (CCTV) por cada dos habitantes. Además, se estima que hoy alguien que hace negocios en Londres es captado por 300 cámaras durante el día y que, en Atlanta, ya hay 16 cámaras de videovigilancia por cada millar de personas (Velasco, 2019).

La pandemia de COVID-19 sin duda aceleró la consolidación de sociedades hipervigiladas por el Estado mediante el uso de tecnologías de inteligencia artificial que permiten procesar volúmenes masivos de datos para hacerle seguimiento automatizado al comportamiento de millones de personas (Harari, 2020). En Rusia, por ejemplo, la videovigilancia con reconocimiento facial ha sido utilizada para verificar el uso del

1 El Carnegie Endowment for International Peace puso a disposición una base de datos con la visualización gráfica de la distribución mundial de las tecnologías de vigilancia que utilizan inteligencia artificial, entre ellas la biometría facial: https://bit.ly/3v5gvyU.

tapabocas por parte de los ciudadanos en el espacio público e imponer multas en caso de contravención (Roussi, 2020, p. 347; Reevell, 2020).

En el último año y medio, los sistemas de monitoreo del desplazamiento mediante aplicaciones móviles y los modelos de videovigilancia del espacio público fueron la punta de lanza de los Gobiernos para rastrear los contagios y controlar la expansión del coronavirus. En América Latina, 16 países implementaron aplicaciones móviles de seguimiento para recabar información en tiempo real con el fin de monitorear y predecir la propagación del virus SARS-CoV-2[2]. La experiencia de estos países con las aplicaciones comparte la tensión que se presenta entre el derecho a la privacidad, el cual se ve afectado en función del manejo que se hace de los datos personales recopilados, y el derecho a la salud pública (Nájera y Ricaurte, 2020, p. 6-8).

El potencial de desarrollo de la biometría facial para mejorar la comprensión de nuestra especie y el funcionamiento de las sociedades humanas está lleno de ambigüedades. Mediante el procesamiento computarizado de las características del rostro hoy se puede inferir con bastante fiabilidad la identidad de las personas, pero hay quienes defienden la posibilidad de deducir incluso rasgos de su personalidad gracias a esta tecnología.

Un investigador de Stanford, Michal Kosinski (2021), sostiene haber diseñado un algoritmo de reconocimiento facial que puede detectar la orientación ideológica liberal o conservadora de los individuos con una efectividad del 72 %, una tasa de éxito significativamente superior a la de la suerte (50 %) y el humano (55 %). Los críticos de Kosinski señalan que sus resultados son la consecuencia de un diseño experimental defectuoso y una mala interpretación de los datos. Además, asocian su trabajo al determinismo de la teoría fisonomista de Cesare Lombroso, el criminólogo italiano de comienzos del siglo XX que creyó descubrir los orígenes biológicos del crimen y poder identificarlos en los rasgos físicos del delincuente habitual: Lombroso incluso publicó una taxonomía criminal en la que afirmaba que "casi todos los criminales" tienen "orejas de jarra, pelo grueso, barbas finas, senos pronunciados, barbilla prominente y pómulos anchos" (Wiggers, 2021).

En la misma línea, algunas iniciativas tecnológicas como "Faception" parecen haber sobrevendido las capacidades de la tecnología de reconocimiento facial, al punto de afirmar que de ella se puede obtener información sobre características

2 En la región, solo El Salvador, Haití, Nicaragua y Venezuela no recurrieron a este tipo de tecnología de seguimiento de la enfermedad.

de la personalidad como el coeficiente intelectual, la capacidad de análisis, la perseverancia y otras habilidades sociales del individuo[3]. En sentido contrario, Feldman *et al*. (2019) argumentan que las expresiones faciales que comunican ira, miedo, felicidad, tristeza y sorpresa "varían sustancialmente entre culturas, situaciones e incluso entre gente en una condición particular", mientras que configuraciones de movimientos faciales similares pueden expresar más de una emoción. En otras palabras, la complejidad de la expresividad facial humana haría, si no inviables, al menos muy difíciles de cumplir esas promesas de las TRF.

La *startup* israelí Faception llega a aseverar en su sitio web que puede detectar "potenciales terroristas o criminales antes de que tengan la oportunidad de hacer daño". Se podría tratar, en este caso, de una versión contemporánea de la frenología, una pseudociencia del siglo XIX que creía poder inferir la proclividad al crimen con base en el tamaño del cráneo y la fisonomía humana (Pascual, 2021). El uso de la TRF con fines de perfilamiento criminal plantea, además, un alto riesgo de judicialización de falsos positivos debido a la imprecisión de los algoritmos, así como de normalización de sesgos algorítmicos injustos que pueden terminar por automatizar los prejuicios ya presentes en nuestras sociedades (Lubin, 2016; O'Neil, 2016). Sobre este peligro profundizaremos en la quinta parte de esta investigación.

También avanzan proyectos de investigación sobre el uso de TRF en otras especies, como los cerdos, dirigidos a lograr una mejor comprensión de la emocionalidad animal gracias al análisis algorítmico de su expresividad facial (Wee y Chen, 2019).

En definitiva, mientras la perspectiva optimista sobre las potencialidades de las TRF causa asombro y esperanza por sus promesas como herramienta que podría mejorar el mundo, una visión escéptica también advierte sobre los preocupantes riesgos de una eventual deriva de nuestras sociedades hacia escenarios orwellianos de supervigilancia arbitraria por parte del Estado y las corporaciones.

Alimentados por ese temor, los grupos activistas que promueven la prohibición de las TRF en el espacio público o su uso restringido se han multiplicado en todas las latitudes. En Reino Unido y Estados Unidos se destacan las organizaciones sin ánimo de lucro Statewatch[4] y Fight for the Future[5]. En América Latina, la iniciativa Reconocimientofacial.info busca "que el re-

3 https://www.faception.com

4 https://www.statewatch.org

5 https://www.fightforthefuture.org

conocimiento facial se utilice en forma excepcional, acotada y proporcionada, y bajo los más altos estándares de control y transparencia"[6]. En Colombia, la Fundación Karisma, que se dedica a la "promoción de los derechos humanos en el mundo digital", ha asumido el liderazgo de esta causa con publicaciones críticas sobre el avance de las TRF en el país[7].

Esta investigación asume una postura que busca evitar tanto el optimismo como el catastrofismo tecnológico, con base en la evidencia disponible, a propósito del análisis de la eventual ampliación de los sistemas de videovigilancia en Colombia y su fortalecimiento con la implementación de *software* de reconocimiento facial automatizado[8]. Como todas las nuevas tecnologías, las TRF plantean riesgos, pero también oportunidades para mejorar nuestras sociedades. Existen usos legítimos de los sistemas de videovigilancia para garantizar la seguridad y prevenir el crimen, mientras se salvaguardan los derechos humanos, se protegen los datos personales y, con ello, se evitan los abusos por parte de las autoridades que los administran.

En lo que concierne a la utilidad de la videovigilancia en materia de seguridad, varios estudios empíricos proveen evidencia de su eficacia en la prevención del crimen (Ratcliffe *et al*., 2009; Welsh y Farrington, 2009). Por ejemplo, en la segunda ciudad colombiana más poblada, Medellín, Gómez *et al.* (2017) encontraron una reducción del 23,5 % en la criminalidad reportada y un declive del 31,5 % en los arrestos registrados dentro de las zonas de nuevo cubrimiento de la videovigilancia, gracias a la instalación de 587 cámaras entre 2013 y 2015.

Otro efecto positivo documentado de la videovigilancia es la disuasión del abuso policial, bien sea mediante la grabación de las intervenciones policiales por las cámaras instaladas en el espacio público, o por la utilización de cámaras corporales (*bodycams*) por los policías, que registran la actividad policial en tiempo real y con ello previenen abusos y discrecionalidad en la actuación (Barbosa *et al*., 2021; Dammert y Silva, 2019, p. 21). Lo anterior, desde luego, siempre que exista control sobre las autoridades que administran el material grabado, para evitar su manipulación y ocultamiento estratégico por parte de las fuerzas policiales (Jasso y Jasso, 2021). Todos estos resultados presentan a la videovigilancia como una herramienta idónea para cumplir los fines estatales de mantenimiento del orden público, garantía de la seguridad ciudadana y prevención del crimen.

6 https://reconocimientofacial.info/por-que-no/

7 Ver por ejemplo Sáenz y Spanger (2018) y https://web.karisma.org.co/camaras-indiscretas/

8 Al momento de escribir estas líneas, el Distrito Capital de Bogotá se encontraba en la etapa de planificación de esta mejora, mediante un proceso de articulación interinstitucional entre la Secretaría de Seguridad y Convivencia Ciudadana, la Registraduría Nacional del Estado Civil y la Policía Nacional.

Sin embargo, desde la orilla opuesta, la principal crítica a la efectividad de la videovigilancia se basa en el argumento del "desplazamiento": según esta visión, la instalación de circuitos cerrados de televisión (CCTV) no disuade del delito, sino que apenas lo desplaza a zonas no vigiladas (Hesseling, 1994; Armitage, 2002; Jasso, 2020). Además, Norris *et al.* (2004) argumentan que el efecto positivo de la videogilancia tiene que ver más con la percepción de seguridad incrementada que genera en el público, por la idea de que "se está haciendo algo en relación con el problema del delito" (Lio, 2015, p. 283), que con los efectos que se puedan probar sobre la tasa de criminalidad.

Debido a lo reciente de la implementación masiva de esta nueva tecnología, aún no tenemos evidencia estadística de la efectividad aumentada de la videovigilancia con reconocimiento facial respecto de la que no lo tiene para prevenir el crimen. Pero sí contamos con evidencia anecdótica de su uso exitoso en labores policiales.

La policía india, por ejemplo, ha utilizado las cámaras con TRF en los estados de Delhi y Telangana para identificar en las calles a miles de niños extraviados y reunirlos nuevamente con sus familias, luego de haber sido víctimas de la criminalidad organizada, que trafica con ellos y los fuerza a trabajar en fábricas, restaurantes, burdeles o como mendigos (Nagaraj, 2020). La policía de Nueva Delhi reportó haber encontrado tres mil niños extraviados en apenas cuatro días gracias a las cámaras con biometría facial (NDTV, 2018).

Aunque es de esperarse un efecto disuasivo incrementado de la videovigilancia por vía de la mayor probabilidad de judicialización del ofensor, que ofrece la posibilidad de identificarlo gracias a la biometría facial, en virtud de que elevaría aún más el riesgo y el costo del comportamiento criminal (Becker, 1968), aún no contamos con estudios que lo prueben. Por esta razón, se necesita realizar evaluaciones de impacto de la implementación de videovigilancia reforzada con reconocimiento facial.

Es palmaria la intervención que cualquier *sistema de videovigilancia con reconocimiento facial* (en adelante, SVRF) supone en los derechos fundamentales de la intimidad y el *habeas data* de los titulares de la información, cuya imagen e identidad son almacenadas y procesadas por las TRF con fines de vigilancia, monitoreo y eventual perfilación criminal. Se trata de una tecnología que suscita una colisión entre principios constitucionales y derechos fundamentales de la máxima jerarquía en el ordenamiento jurídico internacional, regional y colombiano.

Esta investigación asume el reto de aproximarse al marco jurídico deseable para garantizar que la implementación de un

SVRF en Colombia responda a los más altos estándares de protección de los derechos humanos. Para el efecto, el texto se organiza en cinco partes más que tienen los siguientes objetivos: presentar una breve descripción de la geopolítica mundial de la inteligencia artificial y las TRF (2); caracterizar el desarrollo normativo asimétrico (a distintas velocidades) que han tenido la inteligencia artificial y las TRF en distintas regiones y países del mundo (3); identificar los principales riesgos de vulneración de los derechos humanos que supone la implementación de un SVRF (4); definir el marco normativo aplicable en Colombia a los SVRF, que resulta idóneo para mitigar los dos principales riesgos de vulneración de los derechos humanos detectados de la manera más eficaz (5); comunicar algunas conclusiones sobre el presente y el futuro de la videovigilancia con reconocimiento facial en Colombia y Latinoamérica (6).

2. Geopolítica de la inteligencia artificial y el reconocimiento facial

El término *inteligencia artificial* es utilizado para describir "un conjunto de técnicas informáticas que permiten a los sistemas realizar tareas que normalmente requieren de la inteligencia humana, tales como la percepción visual, el reconocimiento de voz, la toma de decisiones y la traducción de idiomas" (The Economist Intelligence Unit, 2016, p. 3).

La *tecnología de reconocimiento facial*, por su parte, es una modalidad de inteligencia artificial "capaz de identificar a una persona o de comprobar su identidad mediante la comparación y el análisis de los modelos, formas y proporciones de sus rasgos y contornos faciales". Para ello, usa "un algoritmo que codifica automáticamente la imagen parcial (sobre la que se quiere indagar) introducida en [un] sistema, para luego compararla con los perfiles almacenados en él. De este modo se obtiene una lista de 'candidatos' de los aciertos más probables" (Interpol, 2020).

Actualmente, la inteligencia artificial (IA) es un elemento de poder en el escenario geopolítico mundial. "La inteligencia artificial es el futuro... Quien sea que se vuelva el líder en esta esfera será el gobernante del mundo", afirmó Vladimir Putin ante un grupo de estudiantes y periodistas en septiembre de 2017. Tres días después, Elon Musk, el fundador de SpaceX y Tesla, quien hoy además es el hombre más rico del planeta, escribió en Twitter que "la competencia por la superioridad en IA al nivel nacional será la causa más probable de la Tercera Guerra Mundial"[9] (Miailhe, 2018).

Sin caer en el catastrofismo de Musk, es importante conocer la geopolítica mundial de la inteligencia artificial (IA) en general (Prakash, 2018), y de la tecnología de reconocimiento facial en particular, para entender las asimetrías territoriales que se presentan en el desarrollo ético y normativo que acompaña su implementación, así como para dimensionar el tamaño del reto regulatorio que enfrenta Latinoamérica en la materia.

Las tecnologías de IA son producidas en su gran mayoría en Estados Unidos, Europa y Asia. Entre 1990 y 2018, mientras el 51 % de las patentes relacionadas con IA se produjeron en América del Norte y el 23 % en Europa y Asia Central, América Latina y el Caribe registraron menos del 1 % de las patentes (Human-Centered Artificial Intelligence Institut of Stanford University, 2019, p. 30). Por su parte, del total mundial de patentes registradas entre 2000 y 2020, el 7,5 % correspondió a tecnologías de AI, de las cuales Norteamérica produjo el 3,2 %; Asia Oriental y el Pacífico, el 2,6 %, y Europa y Asia Central, el 1,6 % (Human-Centered Artificial Intelligence Institut of Stanford University, 2021, p. 181). Esta distribución altamente desigual de la innovación muestra que la región latinoamericana es apenas *receptora* de estas nuevas tecnologías, las cuales son concebidas y diseñadas según la visión, estrategia e intereses de los *centros productores*.

La concentración del desarrollo de la IA en unas cuantas regiones del planeta ha hecho que se conformen "tribus insulares" de individuos con características muy similares que definen sus características. Se trata de personas que viven y trabajan en Norteamérica y China, que se formaron en las mismas universidades, con un nivel educativo alto, adineradas, mayoritariamente hombres y que comparten unas reglas sociales y aspiraciones comunes. La principal consecuencia de este predominio geopolítico de las tribus de creadores de IA es que terminan por priorizar sus valores, ideales y visiones sobre el resto del mundo (Webb, 2020, p. 81-82).

9 https://twitter.com/elonmusk/status/904638455761612800

El dominio de la IA no es solo el privilegio de unos pocos países, sino de un puñado de compañías. Amy Webb (2020, p. 142) denomina a las principales empresas que lideran el desarrollo de la IA en el mundo "los nueve gigantes", que están conformados por el grupo chino BAT (Baidu, Alibaba y Tencent) y la GMAFIA estadounidense (Google, Microsoft, Amazon, Facebook, IBM y Apple). El modelo consumista estadounidense de explotación de la IA responde a la economía de libre mercado, cotiza en la bolsa de Wall Street y su objetivo central es el lucro. De su lado, el modelo centralizado del Gobierno chino obedece las directrices del Partido Comunista de China, que gobierna ese país bajo un modelo autoritario de hipervigilancia social y quiere posicionarse en la próxima década como la primera potencia tecnológica mundial. Ninguno de los dos modelos es intrínsecamente malo, pero tampoco responde prioritariamente a fines altruistas o que le apuesten al bienestar general de la humanidad.

En el campo específico de las TRF, los principales centros de consumo son Estados Unidos, China e India. En la India opera la base de datos biométrica más grande del mundo, el Proyecto Aadhaar, que contiene los datos personales y de identificación biométrica de los cerca de 1300 millones de habitantes que tiene el país (Thales Group, 2021). En orden alfabético y por nacionalidad, algunas de las empresas líderes en el mercado global del reconocimiento facial son Amazon, Animetrics, Aware, Fulcrum Biometrics, Kairos, Nuance Communications, Stereovision Imaging, TrueFace.AI y Microsoft, de Estados Unidos; Ayonix y NEC, de Japón; Thales e Idemia, de Francia; BioID y Cognitec Systems GmbH, de Alemania; FacePhi y Herta Security, de España; la holandesa SightCorp; id3 Technologies, de Israel; nViso, de Suiza; Neurotechnology, de Lituania; Daon, de Irlanda; Techno Brain, de Dubai; Innovatrics, de Eslovaquia (Mehra, 2020).

Esto hace que la producción y consumo de TRF corresponda a la economía política de esas naciones y reproduzca las creencias, prejuicios y sesgos que prevalecen en las poblaciones y gobiernos de esos países. Por ejemplo, para un Estado totalitario como el chino es de esperarse que la prioridad sea utilizar las TRF para incrementar el control social y la vigilancia estatal, mientras se desconocen múltiples derechos de la población en un país que, por lo demás, no reconoce el sistema internacional de protección de los derechos humanos. Algo similar ocurre en Estados Unidos, un país donde existe una persistente cultura de racismo contra la población afroamericana, y en el cual no sorprende que haya sido documentado el sesgo algorítmico injusto de algunos sistemas de reconocimiento facial contra las mujeres de piel oscura, respecto de quienes el

porcentaje de error, al momento de identificarlas, es mayor: 34,7 %, frente a 0,8 % cuando se trata de hombres de piel clara (Buolamwini y Gebru, 2018).

3. Una reacción regulatoria a distintas velocidades

La reacción regulatoria frente a las nuevas tecnologías va ligada a su mejor comprensión y desarrollo técnico. En el Reino Unido, Webster (2004) identificó tres momentos históricos en el proceso de difusión y regulación de la tecnología de videovigilancia que resulta pertinente extrapolar al desarrollo del reconocimiento facial automatizado.

Una primera etapa denominada "era de la innovación", que tuvo lugar en la primera mitad de la década de los noventa en Gran Bretaña y se caracterizó por la ausencia de regulación e información técnica limitada sobre la configuración y funcionamiento de los sistemas. Una segunda etapa, bautizada como "era del despegue o aceptación" (*Era of Uptake*), que ocurrió en la segunda mitad de la década de los noventa cuando los CCTV se expandieron a una gran cantidad de espacios públicos y se produjo una autorregulación voluntaria que coincidió con el nacimiento de redes de políticas públicas alrededor de la tecnología, conformadas por proveedores del servicio, la policía, políticos y hacedores de política pública. En esta época, proliferaron las guías de uso de la tecnología que no eran legalmente vinculantes, así como los procedimientos de auditoría para verificar su buen uso. Por último, la "era de la sofisticación" llegó a finales de los noventa con una corregulación que combinó la legislación jurídica formal con códigos éticos y un discurso político consolidado a favor de la tecnología. En suma, a medida que la tecnología se volvió más sofisticada (en esta época apareció además el *software* de videoanalítica automatizada) también la legislación alcanzó un mayor desarrollo y se volvió más específica.

Con la biometría facial pareciéramos estar asistiendo actualmente a la "etapa del despegue" con la difusión masiva de esta tecnología en la mayoría de países, mientras que la "etapa de la sofisticación" apenas ha comenzado en un puñado

de naciones, y en otras la ausencia de regulación típica de la "etapa de la innovación" es la regla, bien sea porque resulta estratégica para los intereses de Estados autoritarios o porque la tecnología acaba de llegar y el sistema jurídico aún no ha tenido tiempo de reaccionar.

En definitiva, lo que se puede observar es un desarrollo asimétrico de la reacción desde la ética y desde el derecho frente a las oportunidades y riesgos que plantean las TRF en función del contexto y las prioridades de los países donde se implementan.

El desarrollo de un marco ético y normativo eficaz para la mitigación de los riesgos que plantea la IA se ha convertido en una de las prioridades de la comunidad internacional en los últimos años. De hecho, es tal vez el debate ético-jurídico más urgente en la actualidad.

Los marcos éticos que se han producido para garantizar una implementación de la IA que mitigue sus riesgos son numerosos. Para finales de 2019, se registraban más de 90 documentos publicados por los Gobiernos, las empresas y otros actores relevantes sobre los principios de la IA (Human-Centered Artificial Intelligence Institut of Stanford University, 2019). Sin embargo, en esta investigación prescindiremos de su análisis por dos razones. La primera es que ya existen bastantes publicaciones que abordan con suficiencia cuáles serían los fundamentos éticos de una "Good AI Society" (Floridi *et al.*, 2018).

La segunda razón es que coincidimos en que lo más urgente en este momento es definir un marco jurídico que delimite y garantice el cumplimiento de las responsabilidades concretas, tanto retrospectivas como prospectivas, de los actores estatales y privados, frente a las posibles violaciones a los derechos humanos que plantea la implementación de la IA, y eso solo se puede conseguir gracias a la fuerza de coerción que caracteriza al derecho.

Aunque los marcos éticos son una parte importante del esfuerzo de autorregulación que hacen los actores de la IA, así como de la corregulación entre los sectores público y privado que siempre es deseable que acompañe el avance tecnológico, también es cierto que la legislación formal es la que mejor asegura el cumplimiento de las obligaciones adquiridas. Entre otras razones, porque la proliferación de marcos éticos, en ausencia de normas jurídicas que los vuelvan obligatorios, también puede convertirse en un "distractor" que posterga la legislación en beneficio de los intereses de actores privados (Lara, 2020). En un escenario como este, las lagunas jurídicas pueden volverse un multiplicador de las violaciones a los derechos humanos.

En esta parte de la reflexión presentaremos entonces una visión comparada, muy general, de algunas tendencias normativas en materia de regulación de la IA y las TRF. La pregunta que resolvemos es qué respuestas se han dado a nivel internacional, desde el derecho y más allá de la ética, frente a los desafíos planteados por la IA.

Se analizará el modelo particularmente cauteloso y tuitivo de los derechos humanos de la Unión Europea (3.1); la desprotección que caracteriza a Rusia y China, dos grandes potencias donde la tecnología está al servicio de gobiernos autoritarios (3.2); el modelo de protección diferenciada estadounidense, marcado por la diversidad de regímenes regulatorios que el federalismo promueve (3.3); y, por último, la experiencia latinoamericana, donde se encuentran países que han avanzado más que otros en el desarrollo de sus marcos normativos (3.4).

3.1. La Unión Europea

La Unión Europea se ha enfocado en construir un modelo tuitivo y humanista frente a la implementación de tecnologías impulsadas por inteligencia artificial (Ortega, 2021). Un punto de referencia normativa es el Reglamento General de Protección de Datos (RGPD), que entró en vigor el 25 de mayo de 2018. Se trata de una normativa de carácter vinculante y aplicación inmediata para los Estados miembros, incluidas las autoridades y empresas privadas, en materia de tratamiento y gobernanza de datos. Este reglamento se aplica en la Unión Europea como una ley de protección de datos oficial que es superior y prevalece sobre la normativa nacional.

Sin embargo, a pesar de los esfuerzos por parte de órganos como la Comisión Europea, el Comité Europeo de Protección de Datos, el Supervisor Europeo de Protección de Datos e incluso el Tribunal Europeo de Derechos Humanos, la regulación específica de la inteligencia artificial aún no se ha concretado de manera vinculante para todos los Estados parte de la Unión Europea. Esta situación hace que el nivel de penetración y aceptación de las TRF varíe según la legislación local.

Para dar un ejemplo, en España una Resolución del 27 de julio de 2021 de la Agencia Española de Protección de Datos (AEPD) estableció que "no resulta justificado" el uso de datos biométricos para identificar a estudiantes en un examen. Por esa razón, exhortó a la Universidad de la Rioja (UNIR), una de las instituciones educativas españolas que estaba utilizando esta técnica, a que "adopte las medidas correctivas encaminadas a evitar que el tratamiento previsto pueda suponer un posible incumplimiento de la legislación de protección de datos" (Farrenas, 2021).

A continuación, mencionaremos algunas iniciativas de la Unión Europea (UE) para regular la implementación de la inteligencia artificial y las TRF con el fin de garantizar la protección de los derechos humanos.

El *Plan Coordinado de IA*, actualizado en 2021 por la Comisión Europea, estableció las directrices generales que deberían implementar los países parte de la UE, así como los distintos foros para discutir temas de interés, entre ellos la "Alianza de IA Europea"[10], en la cual se hablará sobre la identificación biométrica remota (Comisión Europea, 2021a).

Ya existe además una propuesta de reglamento del Parlamento Europeo y del Consejo, del 21 de abril de 2021, en la que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión, que catalogó como riesgosos algunos sistemas de identificación biométrica remotos (Comisión Europea, 2021b). Esta propuesta, según la Comisión...

> *"Complementa el Derecho de la Unión vigente en materia de no discriminación al establecer requisitos específicos que tienen por objeto reducir al mínimo el riesgo de discriminación algorítmica, en particular en lo tocante al diseño y la calidad de los conjuntos de datos empleados para desarrollar sistemas de IA, los cuales van acompañados de obligaciones referentes a la realización de pruebas, la gestión de riesgos, la documentación y la vigilancia humana durante todo el ciclo de vida de tales sistemas"*
> *(Comisión Europea, 2021a, p. 4)*

Esta propuesta de reglamento prohíbe los sistemas de videovigilancia "en tiempo real" con identificación biométrica en lugares de acceso público, salvo en casos excepcionales, pues considera que el uso de esta tecnología...

> *"Invade especialmente los derechos y libertades de las personas afectadas, en la medida en que puede afectar la vida privada de una gran parte de la población, provocar la sensación de estar bajo una vigilancia constante y disuadir indirectamente a los ciudadanos de ejercer su libertad de reunión y otros derechos fundamentales"*
> *(Comisión Europea, 2021a, p. 25)*

10 https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance

Además, señala que la inmediatez de las consecuencias y las escasas oportunidades para verificar los resultados en relación con el uso de este tipo de tecnologías que operan en tiempo real aumentan el riesgo para los derechos y libertades de las personas afectadas por la aplicación de estos sistemas (Comisión Europea, 2021a, p. 25).

En el marco de esta propuesta de regulación, al advertirse que los datos personales podrían llegar a ser los más afectados por las TRF, se solicitó al Comité Europeo de Protección de Datos (CEPD) un concepto sobre su utilización (Comisión Europea, 2021b, p. 21). Este concepto fue emitido por el CEPD, de manera conjunta con el Supervisor Europeo de Protección de Datos (SEPD), el pasado 21 de junio de 2021. Ambos órganos recomendaron prohibir el uso de inteligencia artificial para el reconocimiento automatizado de características humanas en espacios públicos. El principal argumento de esta recomendación fueron los "extremadamente altos riegos planteados por la identificación biométrica remota", que recaen sobre grupos específicos de la sociedad que podrían ser discriminados injustamente por las TRF (European Data Protection Board & European Data Protection Supervisor, 2021).

Otra normativa relevante a este respecto es el artículo 4 del Reglamento General de Protección de Datos (RGPD) antes mencionado. Esta legislación define los *datos biométricos* como aquellos...

> *"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"*
> *(Parlamento Europeo, Reglamento [UE] 2016/679, art. 4, numeral 14).*

Esta norma señala, además, en su artículo 9 y en relación con el tratamiento de categorías especiales de datos personales, que...

> *"Quedan prohibidos el tratamiento de datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca*

a una persona física, datos relativos a la salud o datos relativos a la vida sexual o la orientación sexual de una persona física"
(Parlamento Europeo, Reglamento [UE] 2016/679, art. 9).

El mismo artículo contempla algunas excepciones taxativas, como el consentimiento explícito (literal a), la necesidad del tratamiento del dato para el cumplimiento de obligaciones y el ejercicio de derechos específicos, así como la protección de intereses vitales del interesado, entre otros.

Además, el Reglamento (UE) 2018/1860, el Reglamento (UE) 2018/1861 y el Reglamento (UE) 2018/1862 del Parlamento Europeo y del Consejo abordaron lo relacionado con el uso del Sistema de Información Schengen (SIS), que recopila datos biométricos en las fronteras del espacio Schengen. Esta regulación resulta relevante porque estableció una hoja de ruta sobre quiénes son los responsables de los datos y la información biométrica obtenida mediante videovigilancia. No obstante, aún quedan vacíos que la Comisión Europea, la CEPD y el SEPD buscan regular.

3.2. El uso autoritario de las TRF en China y Rusia

El temor por los riesgos que rodean a la implementación de la videovigilancia con reconocimiento facial automatizado alcanzó el culmen en los tres últimos años, en buena medida por el mal uso que han hecho de esta tecnología China y Rusia, que constituyen los dos casos emblemáticos de utilización autoritaria de las TRF por parte del Estado.

Lo que está ocurriendo en China con las TRF es una pesadilla orwelliana convertida en realidad, un sueño distópico que está ocurriendo. China tiene la red de cámaras de vigilancia más grande del planeta. Del total de 770 millones de cámaras de vigilancia instaladas en el mundo, el 54 % se encuentran en China. Esto significa que ese país cuenta con aproximadamente 415,8 millones de dispositivos de videovigilancia (Bischoff, 2021).

El autoritarismo chino del futuro sin duda le apuesta a la alta tecnología como principal herramienta de vigilancia y control social. En ese país ya es frecuente el uso de gafas oscuras inteligentes con capacidad de reconocimiento facial por parte de los policías, en el espacio público y en lugares estratégicos como las estaciones de tren, para monitorizar y detectar a inmigrantes ilegales y delincuentes reportados en las bases de datos. También se han instalado pantallas gigantes que pro-

yectan en las calles los rostros y números de identidad de peatones imprudentes y deudores morosos. Se trata de muros de la vergüenza que aumentan la percepción de vigilancia, modelan el comportamiento y materializan así el panóptico del futuro (Mozur, 2018).

Todas estas medidas de vigilancia se insertan en el dantesco sistema nacional de Puntaje de Crédito Social de China, en el cual los ciudadanos pierden puntos de crédito social cuando cometen infracciones de tránsito, no pagan sus deudas o compran cierto tipo de cosas de las que se infiere un comportamiento irresponsable (también ganan puntos cuando ejecutan "actos heroicos") (Webb, 2020, p. 120-127; Botsman, 2017).

La biometría facial también ha facilitado en China la persecución de la minoría musulmana "uighur", para mapear sus relaciones con amigos y familia y encerrarla en campos de detención luego de identificada (Mozur, 2019a). Otro uso indebido de esta tecnología ha sido la perfilación de los opositores con el fin de reprimir las protestas sociales contra el gobierno en Hong Kong (Mozur, 2019b; Doffman, 2019).

En Rusia, la TRF ha sido utilizada para perseguir, judicializar e incluso desaparecer a opositores del gobierno. Además, existe un mercado negro donde se trafica con la identidad y la localización de las personas en Moscú (Reevell, 2020). Según Amnistía Internacional (2020), Rusia está utilizando tecnología invasiva de reconocimiento facial para reprimir protestas pacíficas contra el gobierno. En palabras de Natalia Zviagina, directora de Amnistía Internacional Rusia:

> *"La tecnología de reconocimiento facial es, por naturaleza, enormemente invasiva, pues permite la vigilancia, recogida, almacenamiento y análisis generalizados y masivos de datos personales confidenciales sin que medie una sospecha razonable individualizada. En manos de las autoridades rusas, que ya perpetran numerosos abusos, y ante la ausencia total de transparencia y rendición de cuentas de estos sistemas, es una herramienta que probablemente se usará para llevar las represalias contra protestas pacíficas a un nivel totalmente nuevo. Es significativo que el gobierno ruso no haya ofrecido explicación alguna sobre la forma en que garantizará el derecho a la intimidad y otros derechos humanos, y que tampoco haya abordado la necesidad de supervisión pública de unas tecnologías tan potentes" (Amnistía Internacional, 2020).*

La tecnología china de reconocimiento facial fabricada por Huawei llegó en 2020 a Serbia, país europeo que ya instaló un millar de esas cámaras en su capital, Belgrado, sin un marco legal claro para su funcionamiento, y planea instalar ocho mil más, mientras los activistas de derechos humanos se preocupan de que esa joven democracia vuelva a prácticas de su pasado autoritario (Roussi, 2020, p. 351).

El paradigma chino se replica en otros países asiáticos con gobiernos autoritarios como Corea del Norte, régimen del que se afirma que podría estar vendiendo en secreto tecnología de reconocimiento facial y de espionaje en el extranjero a través de compañías fachada (Fussell, 2018).

3.3. Estados Unidos

En Estados Unidos, el 59 % de la población considera aceptable el uso de la videovigilancia con reconocimiento facial en lugares públicos para neutralizar amenazas a la seguridad y el 56 % confía en que las agencias de seguridad hacen un uso responsable de esa tecnología (Smith, 2019).

Sin embargo, en noviembre de 2020, 11 ciudades estadounidenses habían prohibido el reconocimiento facial automatizado por parte de autoridades en el espacio público (Castelvecchi, 2020, p. 347). San Francisco, la ciudad más tecnológica del mundo, fue la primera en hacerlo el 14 de mayo de 2019, por considerar que debía primar el derecho a la privacidad de las personas, que las TRF eran un paso hacia una mayor represión estatal cuyos beneficios no están probados. La prohibición también se hizo con el fin de proteger a las minorías de piel más oscura, a quienes los algoritmos de reconocimiento facial identifican con mayor dificultad y pueden conducir a arrestos arbitrarios (Mendoza, 2019) propiciados por una "automatización de los prejuicios" (Eubanks, 2018 y 2019).

El 9 de julio de 2021 entró en vigencia en la Ciudad de Nueva York la "Ley de Información de Identificadores Biométricos". Esta ley les prohibió a los negocios en los que funcione un "lugar de entretenimiento, una tienda minorista o un establecimiento de comida y bebida" vender o intercambiar información de identificación biométrica de cualquier individuo, incluidos sus empleados. Por su parte, la recopilación, uso y retención de los datos biométricos solo se autorizó si se les advierte a los clientes de ella mediante un aviso "en lenguaje claro y sencillo". Otra particularidad de esta ley es que creó un derecho de acción privado, aplicable a "cualquier persona que sea agraviada por una infracción", que les permite a los consumidores de la Ciudad de Nueva York (sin necesidad de que sean

residentes de la ciudad) iniciar una acción judicial, después de darle un período de saneamiento de 30 días al negocio para cumplir con el requisito de notificación.

Estados Unidos ha desarrollado la videovigilancia a partir de la Cuarta Enmienda de la Constitución, la cual protege como principio general el derecho a la privacidad y a que esta no sufra invasiones arbitrarias. Bajo este marco constitucional, la Corte Suprema ha establecido que la "expectativa razonable y legítima" de privacidad de los ciudadanos varía en función de los espacios en que se produce la videovigilancia, pues los espacios públicos son más susceptibles de ella (Williams, 2019).

El 1 de enero de 2021 fue aprobada la Ley Nacional de Inteligencia Artificial en Estados Unidos, una legislación que busca coordinar los esfuerzos de todo el Gobierno federal para acelerar la investigación y la aplicación de la IA como motor de la prosperidad económica y la seguridad nacional de país. Otro objetivo declarado de esta ley es garantizar el liderazgo mundial de EE. UU. en el desarrollo y el uso de IA confiable en los sectores público y privado, además de preparar a la fuerza laboral estadounidense presente y futura para la integración con los sistemas de IA en todos los sectores de la economía y la sociedad.

Una de las grandes virtudes de esta ley es que acogió una definición clara, sobria y muy concreta del término *inteligencia artificial*:

> *"Sistema basado en máquinas que puede, para un conjunto dado de objetivos definidos por el humano, hacer predicciones, recomendaciones o tomar decisiones que influyen en entornos reales o virtuales. Los sistemas de inteligencia artificial utilizan insumos derivados de máquinas y humanos para:*
>
> ***(A)** percibir entornos reales y virtuales;*
>
> ***(B)** abstraer tales percepciones en modelos a través de análisis automatizados; y*
>
> ***(C)** utilizar modelos de inferencia para formular opciones de información o acción".*

3.4. Latinoamérica

En América Latina, a diferencia de Europa, no tenemos unas bases supranacionales comunes en relación con la implementación de tecnologías de inteligencia artificial. Sin embargo, alguna normativa del Sistema Interamericano de Protección de Derechos Humanos (en adelante, SIDH) puede aplicarse para garantizar la efectividad de los derechos humanos frente a las tecnologías emergentes.

Derechos que podrían verse vulnerados por las TRF como el derecho a la no discriminación, el deber del Estado de no injerir en el desarrollo de los derechos y libertades de las personas, el derecho a la intimidad, los derechos políticos (entre otros que serán analizados en la cuarta parte) son protegidos por el SIDH.

La reacción normativa y la producción de política pública en América Latina frente a los retos regulatorios que plantea la inteligencia artificial ha variado según el país. Un trabajo del Banco Interamericano de Desarrollo y C Minds (2020) evaluó el avance en doce países de la región relativo a los niveles de desarrollo en la materia por sectores de la sociedad: gobierno, academia, ecosistema de emprendimiento y sociedad civil. Los porcentajes de avance en cada uno de estos cuatro sectores se calcularon de la siguiente manera: para "gobierno" se usó el promedio del Índice de Adopción Digital del Banco Mundial, el Índice de Disponibilidad de Red del Foro Económico Mundial (FEM) y el índice de formulación de estrategias digital, de datos y de IA (en este último ítem se otorgó un porcentaje de 100 % si el país ya cuenta con ellas, 50 % si están en curso o en espera y 0 % si no las tiene). Para "academia", la cifra se obtuvo de calcular el porcentaje de universidades públicas, no centralizadas, con carreras afines a la IA, que tienen investigación relacionada con la IA y con laboratorio de IA, dividido por el total de universidades estudiadas (3, 4 o 5). Para el sector "ecosistema de emprendimiento" se tomó el indicador de contexto nacional de emprendimiento (NECI), que mide lo favorable que es el entorno para los emprendedores. Para "sociedad civil" se usó el indicador de Civicus Monitor, el cual mide el estado de las libertades de la sociedad civil (Tabla 1).

Tabla 1. Panorámica de los avances en la IA en 12 países seleccionados del ALC (%)

	Gobierno	Academia	Ecosistema emprendimiento	Sociedad Civil
Argentina	77,5	52	52,4	75
Brasil	77,4	100	41,8	50
Chile	79,1	74,6	50,9	75
Colombia	89,1	68	47,3	25
Costa Rica	65,1	75	n/d	100
Ecuador	61,4	60	n/d	75
México	76,4	64	52,1	25
Paraguay	61,8	46,4	n/d	50
Perú	62,8	46,4	45,4	50
República Dominicana	61,6	46,4	45,7	75
Trinidad y Tobago	38,5	60,3	n/d	75
Uruguay	91,5	60	47,2	100

Fuente: Banco Interamericano de Desarrollo y C Minds (2020).

Algo similar ocurre con el desarrollo y aceptación en la región de las tecnologías de biometría. El trabajo de la Asociación por los Derechos Civiles (2017) sobre cómo las sociedades latinoamericanas se enfrentan a la implementación de las tecnologías biométricas reveló distintos niveles de penetración de estas tecnologías en la prestación de servicios públicos y privados. Los cuatro países de la región que más registraron avances con los sistemas de identificación multibiométrica (dactilar, facial y de iris) fueron Argentina, Colombia, Chile y México.

4.
Principales riesgos de la videovigilancia con reconocimiento facial en materia de derechos humanos

La urgencia global de una "agenda internacional de derechos humanos" en materia de inteligencia artificial ha sido diagnosticada por varios estudios (Risse, 2018 y 2019; Donahoe y MacDuffe, 2019). Esta agenda requeriría incluso de la suscripción de un nuevo tratado internacional de derechos humanos específico para estas tecnologías que sea vinculante para los Estados y las empresas (Lara, 2020, pp. 52 y 60).

Sin embargo, mientras se alcanza ese grado de desarrollo regulatorio global, lo más pragmático es aplicar los tratados internacionales de derechos humanos ya existentes que permiten enmarcar una implementación segura de la IA. Esta estrategia tiene la doble ventaja de que las violaciones que se cometan no solo hacen responsables a los Estados, sino también a las organizaciones privadas, en virtud de los Principios Rectores sobre las Empresas y los Derechos Humanos de Naciones Unidas, (Rasso *et al.*, 2018, p. 12), en particular el principio 13, que las obliga a que...

> *"a) Eviten que sus propias actividades provoquen o contribuyan a provocar consecuencias negativas sobre los derechos humanos y hagan frente a esas consecuencias cuando se produzcan; b) Traten de prevenir o mitigar las consecuencias negativas sobre los derechos humanos directamente relacionadas con operaciones, productos o servicios prestados por sus relaciones comerciales, incluso cuando no hayan contribuido a generarlos" (Naciones Unidas, 2011).*

El reto más inmediato consiste entonces en identificar los principales derechos que se encuentran bajo amenaza por la IA y la TRF, para definir las estrategias de mitigación de riesgos. Para este propósito, seguiremos la metodología de identificación de niveles de riesgo por cada derecho, propuesta por Raso *et al.* (2018). El análisis del riesgo de vulneración de los derechos humanos por los SVRF se hará respecto de los derechos contenidos en tres instrumentos internacionales y que revisten relevancia para las TRF: la Declaración Universal

de Derechos Humanos (DUDH), la Convención Americana sobre Derechos Humanos (CADH) y el Convenio Europeo de Derechos Humanos (CEDH). Con base en esta última, Wagner (2018) efectuó un ejercicio similar para el contexto europeo.

El nivel de riesgo que corre cada derecho se especifica en la Tabla 2. Allí se muestra si el impacto de la tecnología resulta potencialmente positivo o negativo[11]. El posible impacto, tanto positivo como negativo, se evaluará en una escala tripartita de riesgo o beneficio *bajo*, *medio* o *alto*.

Tabla 2. Efectos de las TRF sobre los derechos humanos

Derecho		Instrumento	Impacto sobre los DH	
			Positivo	Negativo
1.	Derecho a la vida, libertad y seguridad personal	CADH (Arts. 4 y 7) CEDH (Arts. 2 y 5) DUDH(Art. 3)	Alto	
2.	Derecho a la libre circulación y residencia	CADH (Art. 22) DUDH(Art. 13)		Bajo
3.	Derecho a la igualdad	CADH (Art. 24) DUDH(Art. 21)		Alto
4.	Derecho a la no discriminación	CADH (Art. 1) CEDH (Art. 14) DUDH(Art. 7)		Alto
5.	Derecho a la igualdad y protección ante la ley contra cualquier discriminación	CADH (Art. 24) DUDH(Art. 7)		Medio
6.	Deber del estado de no injerir en el desarrollo de los derechos y libertades de las personas	CADH (Art. 1)		Medio
7.	Derecho a la intimidad y prohibición de ataques contra la honra y la reputación	CADH (Art. 11) DUDH(Art. 12)		Alto
8.	Prohibición de ser arbitrariamente detenido, preso o desterrado	CADH (Art. 7) CEDH (Art. 5) DUDH(Art. 9)		Alto

CADH: Convención Americana sobre derechos Humanos
CEDH: Convenio Europeo de Derechos Humanos
DUDH: Declaración Universal de los Derechos Humanos

11 Los derechos humanos respecto de los cuales el efecto de las TRF es nulo fueron excluidos de la tabla. Estos derechos son: prohibición de esclavitud, derecho a la integridad personal, derecho a la personalidad jurídica, derecho a la protección judicial, garantías judiciales, derecho de asilo, derecho a la nacionalidad, derecho al matrimonio, derecho a la propiedad, derecho a la libertad de pensamiento, conciencia y religión, derecho a la libertad de opinión y expresión, derecho de reunión y asociación, derecho a la seguridad social, derecho al trabajo y su libre escogencia, derecho al descanso, derecho a un nivel de vida adecuado, derecho a la educación, derecho a la vida cultural en comunidad, deberes esenciales, derechos del niño, derecho de rectificación o respuesta, derecho al desarrollo progresivo y derecho a protección internacional.

9.	Derecho a un jucio público, presunción de inocencia	CADH (Art. 8) CEDH (Art. 6) DUDH(Art. 11)		Alto
10.	Derechos políticos	CADH (Art. 23)		Alto

Alto Medio Bajo

Fuente: elaboración propia.

El diagnóstico de peligrosidad que presenta la Tabla 2 arroja que existen al menos seis derechos humanos bajo riesgo alto de vulneración por parte de las TRF: a la igualdad, a la no discriminación, a la intimidad, a la prohibición de ser detenido arbitrariamente, a un juicio público y a la presunción de inocencia, así como a los derechos políticos, en particular el derecho a la protesta, que se ve amenazado cuando los gobiernos utilizan la videovigilancia con fines represivos y de perfilación de sus opositores, como se vio que ocurre en Rusia y China.

Por su parte, el impacto de las TRF sobre el derecho a la vida, la libertad y la seguridad personal lo consideramos potencialmente positivo, siempre que se mitigue el riesgo de sesgo algorítmico, que podría conducir a restricciones de la libertad o judicializaciones injustas, como consecuencia del perfilamiento o identificación erróneo de potenciales delincuentes.

En el siguiente acápite nos concentraremos en definir el marco normativo idóneo en Colombia para mitigar dos de los principales riesgos altos de vulneración de los derechos humanos identificados. De un lado, el derecho a la intimidad, en particular, en dos de sus manifestaciones que son el derecho a la protección de datos personales y el *habeas data*, que pueden verse afectados cuando los datos biométricos no son tratados con la confidencialidad que exige su carácter de datos sensibles. De otro lado, el derecho a la igualdad y la no discriminación, que se ve amenazado por la eventual aplicación de sesgos algorítmicos injustos que perjudican a grupos específicos de la sociedad.

5. Marco normativo de la inteligencia artificial y la videovigilancia con reconocimiento facial en Colombia frente a los dos principales riesgos que plantean para los derechos humanos

En Colombia no existe aún una ley específica de inteligencia artificial ni de videovigilancia con reconocimiento facial. Sin embargo, hay bastante normativa aplicable a varios de los aspectos que involucra el despliegue de esta tecnología, además de los tratados internacionales sobre derechos humanos aprobados por el Congreso.

Existen dos documentos de política pública de relevancia en Colombia que se ocupan de definir la estrategia nacional en materia de explotación de datos e inteligencia artificial. El Consejo Nacional de Política Económica y Social (CONPES), adscrito al Departamento Nacional de Planeación, emitió dos documentos CONPES. El primero, la Política Nacional de Explotación de Datos (Big Data), que tiene por objetivo aumentar el aprovechamiento de datos mediante el desarrollo de las condiciones para que sean gestionados como activos para generar valor social y económico. En lo que se refiere a las actividades de las entidades públicas, esta generación de valor es entendida como la provisión de bienes públicos para brindar respuestas efectivas y útiles frente a las necesidades sociales (Consejo Nacional de Política Económica y Social, 2018).

El segundo documento CONPES, la Política Nacional para la Transformación Digital e Inteligencia Artificial, busca la transformación digital del Estado, porque considera que este fenómeno está cambiando radicalmente la sociedad y es uno de los principales motores de la Cuarta Revolución Industrial (4RI). Esta transformación plantea grandes retos para el país, que en caso de no adaptarse rápidamente al nuevo entorno productivo perdería la oportunidad de producir valor económico

y social mediante el uso de las tecnologías de la información y las comunicaciones (TIC) (Consejo Nacional de Política Económica y Social, 2019).

Además, el Gobierno nacional expidió el documento *Modelo de gobernanza de la infraestructura de datos para el desarrollo de tecnologías emergentes*, en el cual señala la necesidad de crear una infraestructura de datos robusta debido a su importancia en la generación de tecnologías emergentes, como la inteligencia artificial. Allí se sostiene que los datos se convirtieron en un activo estratégico para cualquier país, que Colombia ha venido realizando esfuerzos importantes para su desarrollo y fácil acceso y que, dada su importancia transversal, es necesario proponer un marco específico para la creación de una infraestructura de datos que cumpla las siguientes características: que sea robusta, que facilite la interoperabilidad, que sea útil para el desarrollo y diseño de las tecnologías emergentes y que tenga el propósito de servir al bien común y al desarrollo sostenible del país (Guío Español, 2021).

En 2021 el Gobierno colombiano también expidió el "Marco Ético para la Inteligencia artificial en Colombia", un documento de *softlaw*, que no es de obligatorio cumplimiento, pero hace recomendaciones basadas en las buenas prácticas internacionales a las entidades públicas para la formulación y gestión de los proyectos que utilicen inteligencia artificial. El principal objetivo de este proyecto es que las entidades del Estado sigan los principios allí consignados en el desarrollo de sus sistemas de IA y publiquen sus características para brindar transparencia y construir confianza entre la ciudadanía (Guío Español *et al*., 2021).

Recientemente, en el Congreso colombiano se tramitó el Proyecto de Ley 021 de 2020, que tenía por objeto establecer los lineamientos de política pública para el desarrollo, uso e implementación de la inteligencia artificial. Esta iniciativa fue impulsada por el Partido Liberal, el Centro Democrático, el Partido de la U, Cambio Radical y Colombia Justa Libres. Sin embargo, el proyecto de ley fue retirado el 3 de junio de 2021 por el autor.

A continuación, expondremos la legislación aplicable en Colombia a la videovigilancia con TRF, que permite mitigar el riesgo de vulneración de los dos derechos más amenazados: el derecho a la intimidad, en sus dos derivaciones del derecho a la protección de datos y el *habeas data*, así como el derecho a la no discriminación por la eventual aplicación de sesgos algorítmicos injustos.

5.1. Derecho a la intimidad, protección de datos y *habeas data*

Para comenzar, la Ley Estatutaria 1581 de 2012 reguló lo relativo a la protección de datos personales. En su artículo 5, esta ley definió la categoría de "datos sensibles" como sigue:

> *"Para los propósitos de la presente ley, se entiende por datos sensibles aquellos que afectan la intimidad del Titular o cuyo uso indebido puede generar su discriminación, tales como aquellos que revelen el origen racial o étnico, la orientación política, las convicciones religiosas o filosóficas, la pertenencia a sindicatos, organizaciones sociales, de derechos humanos o que promueva intereses de cualquier partido político o que garanticen los derechos y garantías de partidos políticos de oposición así como los datos relativos a la salud, a la vida sexual y los* ***datos biométricos****" (negrita añadida).*

El hecho de que sean caracterizados como datos "sensibles" y "privados" por la nomenclatura legal hace que los datos biométricos gocen de protección reforzada en Colombia, en virtud de que los rige el "principio de acceso y circulación restringida" a los titulares de los datos o a los terceros especialmente autorizados por la ley (art. 4 de las leyes 1581 de 2012 y 1266 de 2008).

Adicionalmente, la condición de datos sensibles de los datos biométricos hace que sigan la regla general de prohibición de su tratamiento, salvo que se trate de las cinco excepciones expresamente previstas en el artículo 6 de la Ley 1581 de 2012, a saber:

1. Que el titular haya dado su autorización explícita, salvo en los casos que por ley no sea requerido el otorgamiento de dicha autorización.
2. Que el tratamiento sea necesario para salvaguardar el interés vital del titular y este se encuentre física o jurídicamente incapacitado, caso en el cual los representantes legales deberán otorgar su autorización.
3. Que el tratamiento se haga en el curso de las actividades legítimas y con las debidas garantías por parte de una fundación, ONG, asociación o cualquier otro organismo sin ánimo de lucro, cuya finalidad sea política, filosófica, religiosa o sindical, siempre que se refieran exclusivamente a sus miembros o a las personas que mantengan contactos regulares por razón de su finalidad. En estos casos, los da-

tos no se podrán suministrar a terceros sin la autorización del titular.

4. Que el tratamiento se refiera a datos que sean necesarios para el reconocimiento, ejercicio o defensa de un derecho en un proceso judicial.

5. Que el tratamiento tenga una finalidad histórica, estadística o científica, caso en el cual deberán anonimizarse los datos, es decir, garantizarse la supresión de la identidad de sus titulares.

Otro elemento regulatorio importante es la guía denominada *Protección de datos personales en sistemas de videovigilancia*, publicada por la Superintendencia de Industria y Comercio (2016), la autoridad pública colombiana que, a través de la Delegatura de Protección de Datos Personales, es la encargada de vigilar y garantizar que las entidades públicas y privadas respeten los principios, derechos y garantías de tratamiento de los datos personales regulados en la Ley 1266 de 2008 y la Ley 1581 de 2012.

En esta guía se precisó que los sistemas de videovigilancia son considerados como *intrusivos de la privacidad* al involucrar herramientas como el monitoreo y la observación de las actividades que realizan las personas a lo largo del día. Por esta razón, la guía aconseja que antes de tomar la decisión de implementar tales sistemas, se debe tener en cuenta la necesidad de utilizarlos y considerar si esa necesidad solo se suple con su implementación, o si existen otros mecanismos que se podrían utilizar con un menor impacto en la privacidad de las personas. En otras palabras, esta guía adopta el principio de proporcionalidad en la implementación de la videovigilancia.

Por su parte, la jurisprudencia de la Corte Constitucional colombiana también se ha pronunciado sobre la videovigilancia. En la Sentencia T-114 de 2018, este tribunal señaló que...

> *"Precisamente, la Superintendencia de Industria y Comercio publicó la guía denominada "Protección de datos personales en sistemas de videovigilancia", en la cual se brinda una orientación a aquellas personas naturales o jurídicas que implementen tales sistemas y, en consecuencia, los exhorta para que adecúen el uso de los mismos a las disposiciones que regulan la protección de datos personales".*

La jurisprudencia de esta Corte también ha reconocido que los métodos de vigilancia son instrumentos encaminados a la prevención del delito por medio de la disuasión, así como a la identificación de delincuentes en un entorno físico determinado. En consecuencia, el alto tribunal considera que las cámaras de seguridad reducen la probabilidad de que se cometan delitos en los espacios bajo vigilancia, porque en ellos resulta más difícil la perpetración impune de una conducta punible (Sentencia T-114 de 2018).

En suma, sobre los alcances y límites constitucionales de la videovigilancia, la Corte Constitucional colombiana se ha pronunciado en varias providencias, entre ellas: la Sentencia T-768 de 2008, sobre las cámaras de vigilancia en el lugar de trabajo; la Sentencia T-407 de 2012, sobre las cámaras de vigilancia en escuelas públicas; las Sentencias T-487 de 2017 y T-114 de 2018, sobre las cámaras de vigilancia en establecimientos de comercio.

En su más reciente jurisprudencia sobre el tema, la Corte Constitucional hizo el control de constitucionalidad de algunos apartes del nuevo Código de Policía y estableció los lineamientos generales para la operación de los sistemas de videovigilancia en espacios públicos y semiprivados, a saber:

> *"1) La autorización de captación y almacenamiento de aquello que ocurre en espacios públicos o semipúblicos resulta conducente para contribuir con la finalidad de darle prevalencia al interés general y la garantía del orden público, permitiendo enfrentar e investigar la comisión de crímenes y las alteraciones de la convivencia; 2) en los espacios públicos y semipúblicos a los que se refiere la norma, se permiten mayores interferencias frente a los derechos; 3) en estos espacios la expectativa de privacidad se reduce; se debe optar en este caso por dar aplicación al principio de armonización concreta, con el propósito de guardar la integridad de la Constitución y, al mismo tiempo, reconocer las competencias del legislador a efectos de regular una materia especialmente compleja" (Sentencia C-094 de 2020).*

Como se ve, jurídicamente la protección de datos personales y, en específico, la confidencialidad de los datos biométricos de los colombianos está garantizada por la normativa hoy vigente en el país. El tratamiento de datos biométricos está sujeto al régimen de datos sensibles, que sigue el principio de confidencialidad y circulación restringida, salvo en las hipótesis de expresa autorización legal.

5.2. Sesgo algorítmico injusto

El riesgo de aplicación de sesgos algorítmicos injustos por parte de las tecnologías de IA en los sectores público y privado ha sido ampliamente documentado (O'Neil, 2016; Eubanks, 2018). Esto ocurre esencialmente porque los algoritmos que sustituyen a la voluntad humana han sido diseñados también por humanos, que definen sus criterios decisionales con base en sus propios prejuicios, de los cuales con frecuencia no son conscientes porque el contexto cultural donde se concibe la IA los ha normalizado (Flórez, 2020).

Tres ejemplos de aplicaciones famosas permiten ilustrar esta problemática en el sector privado. La aplicación Waze escoge por lo general el camino más rápido, pero no le importa el consumo de combustible —por eso a veces escoge trayectos mucho más largos y por ende menos ecológicos— ni la seguridad de la ruta, al punto de que ya varias personas perdieron la vida por seguir una vía peligrosa sugerida por la aplicación (Cortés, 2018). Netflix recomienda producciones muy parecidas a las vistas previamente y, con ello, genera una *filter bubble* (o caja de resonancia) que le impide al televidente innovar con géneros y preferencias distintas. Y Rappi, la aplicación de domicilios, induce el consumo de comidas similares a las anteriormente pedidas por el comensal, con lo cual una persona obesa es condenada por el algoritmo a los malos hábitos alimenticios (Flórez, 2020).

Amazon se vio obligado a desmontar en 2018 un *software* de selección de personal que discriminaba a las mujeres programadoras (Dastin, 2018). Esto sucedió porque el algoritmo había sido entrenado con datos de la última década, en la cual las mujeres tuvieron una presencia mínima en el mercado laboral de los ingenieros de sistemas. Las mujeres tienen hoy una presencia de apenas el 15 % entre los científicos de datos (Duranton *et al.*, 2020). Esta falta de diversidad agrava el problema de los sesgos sexistas que presentan los algoritmos de inteligencia artificial porque impide que las mujeres afectadas participen en su diseño.

En el sector público se reproduce la misma problemática. En Estados Unidos, el sistema judicial se alimenta de varios *softwares* que predicen el riesgo de reincidencia criminal y son usados para informar las decisiones de fiscales y jueces sobre la libertad de las personas. Una auditoría hecha por la organización sin ánimo de lucro ProPublica (Angwin *et al.*, 2016) detectó que el programa perfilaba más drásticamente el riesgo de los afroamericanos que el del resto de la población. En el condado de Broward, Florida, fueron analizados los puntajes

asignados a más de siete mil personas arrestadas. El análisis encontró que los afroamericanos tuvieron una probabilidad 77 % mayor de ser considerados de alto riesgo de cometer un crimen violento futuro y 45 % más grande de incurrir en un crimen de cualquier tipo. En este caso, los derechos a la presunción de inocencia, a un juicio justo y a la igualdad se vieron vulnerados por el uso de una tecnología que no fue suficientemente auditada antes de su implementación.

Como se vio en líneas anteriores, el sesgo algorítmico en contra de la población femenina con piel oscura de algunas TRF utilizadas en Estados Unidos fue puesto en evidencia por Buolamwini y Gebru (2018). Esta importante investigación visibilizó un problema de la tecnología que, por fortuna, se ha probado en los tres últimos años que es susceptible de una solución técnica.

Desde entonces, las empresas desarrolladoras de *software* de reconocimiento facial han reducido el margen de error a magnitudes cercanas a cero y, en todo caso, han ubicado a las TRF por encima de las capacidades humanas de identificación de otros humanos. El *software* de reconocimiento facial LFIS (Live Face Identification Software) registra una tasa éxito en la identificación del 99,44 % (Thales Group, 2021).

Por su parte, el Instituto Nacional de Estándares y Tecnología (NIST, por sus siglas en inglés), una agencia estatal del Departamento de Comercio de EE. UU., encontró que para enero de 2020 los mejores algoritmos ya no presentaban sesgo racial ni de género, ni siquiera cuando las personas usaban tapabocas (Grother *et al*., 2021). "Los críticos estaban equivocados", sostiene un análisis de ese informe (McLaughlin y Castro, 2020).

Durante la última década hemos asistido a un mejoramiento acelerado del reconocimiento facial automatizado: las pruebas de precisión muestran que las TRF están mejorando todo el tiempo su desempeño. Un estudio anterior del NIST (Grother *et al*., 2018) había registrado avances sustanciales en la precisión del reconocimiento entre 2013 y 2018, periodo que superó los mejores récords del periodo 2010-2013: la mayoría de los algoritmos de reconocimiento facial de 2018 superaron al algoritmo más preciso de finales de 2013.

En sus pruebas de 2018, el NIST cuantificó que apenas el 0,2 % de las búsquedas, en una base de datos de 26,6 millones de fotos, no coincidían con la imagen correcta, en comparación con una tasa de fallos del 4 % que se registraba en 2014 (Grother *et al*., 2018). Dos años después, en las pruebas realizadas en 2020 nuevamente por el NIST, el mejor algoritmo de identificación facial tuvo una tasa de error del 0,08 %, es decir, menos de un error por cada 1000 imágenes. Esto significa una mejora de 50

veces en seis años, gracias al desarrollo de los algoritmos de redes neuronales artificiales, que están ayudando a que los algoritmos de reconocimiento facial se vuelvan más precisos (Crumpler, 2020).

También se han multiplicado y sofisticado los escenarios y las metodologías de auditoría de la precisión de los algoritmos de reconocimiento facial (Raji *et al*., 2020). Incluso, Raz *et al*. (2021) desarrollaron un demo interactivo que permite revelar las funciones algorítmicas detrás de las TRF y facilitarles la comprensión de los riesgos a los hacedores de política pública y el público en general.

Los principios de transparencia, responsabilidad, explicabilidad e inteligibilidad algorítmica se encuentran en proceso de gestación y hoy encuentran desarrollo en derechos humanos específicos como el "derecho a obtener explicaciones sobre decisiones individuales automatizadas" (*right to explanation*) (Castaño, 2020). Aunque este derecho aún no ha sido reconocido específicamente en un tratado internacional de derechos humanos, se deduce del derecho a la no discriminación (que sí está previsto tanto en la DUDH como en la CADH y la CEDH) porque la imposibilidad de obtener explicaciones sobre las operaciones algorítmicas puede conducir a discriminaciones injustas.

Además, el principio de transparencia algorítmica ha sido desarrollado en algunas legislaciones nacionales de avanzada en la materia, como la neozelandesa, que aprobó en julio de 2020 la Carta de Algoritmos para Aotearoa de Nueva Zelanda. Esta carta es la primera legislación de su especie en el mundo y es una guía para que las agencias del Gobierno en Nueva Zelanda, que utilizan algoritmos automatizados en la prestación de servicios públicos, lo hagan garantizando la transparencia y la no discriminación en el tratamiento de los datos de los ciudadanos (The Technolawgist, 2020).

Entre los compromisos adquiridos en esta carta por las entidades estatales neozelandesas firmantes se encuentran: dar una explicación de cómo los algoritmos toman las decisiones; asegurarse de que los datos sean adecuados para la finalidad buscada mediante la identificación y gestión de sesgos injustos; garantizar que la privacidad, la ética y los derechos humanos sean protegidos mediante una revisión periódica de los algoritmos por pares; incorporar una perspectiva Te Ao Māori, de acuerdo con el Tratado de Waitangi, en el desarrollo y uso de los algoritmos; y explicar claramente el papel que juegan los humanos en las decisiones informadas por algoritmos (Chanthadavong, 2020).

No se espera que las decisiones que toma la inteligencia artifi-

cial sean perfectas, pero sí que al menos sean mejores que las del humano, gracias a su mayor capacidad de procesamiento de volúmenes masivos de información. La precisión del reconocimiento facial en los humanos ha sido estimada en 97,53 % (The Physics arXiv Blog, 2014) y ya existen algoritmos que en condiciones ideales de captura de la imagen superan el 99 % de precisión (Crumpler, 2020). Esto demuestra que los algoritmos artificiales tienen el potencial de superar los instintos y la intuición en que se basan muchas decisiones humanas, siempre que se haga un control riguroso para que no reproduzcan los sesgos injustos de sus creadores. El desafío consiste en mitigar los tres principales riesgos que se presentan en la implementación de algoritmos de IA: la "opacidad", que hace difícil establecer la justicia de los modelos matemáticos; la "injusticia", que ocurre cuando el algoritmo opera en detrimento de los intereses de grupos específicos de personas; y la "escala" masiva de aplicación del algoritmo, que cuando afecta a enormes cantidades de personas no permite que nadie escape a su operación (O'Neil, 2016; Collins, 2017).

Como estrategia política y jurídica urgente se necesita construir una agenda de derechos humanos en relación con la inteligencia artificial (Risse, 2018). El reto no es promover una postura hostil al desarrollo tecnológico, sino evitar su avance sin una estrategia de mitigación del riesgo que termine por multiplicar los casos de "estupidez artificial" masificada (los cuales, como se vio, también existen).

El primer paso en esta dirección debe ser la ampliación del marco general de derechos que establece el sistema internacional de derechos humanos actual, para así prevenir las amenazas que plantea la inteligencia artificial a la dignidad humana, la responsabilidad democrática, la privacidad, el acceso a la información, la igualdad y el derecho a la no discriminación.

Un segundo paso será generalizar la política de *derechos humanos por diseño (humans-rights-by-design)* (Allison-Hope, 2017). Este enfoque exige que los creadores de cualquier nueva tecnología de inteligencia artificial se comprometan a que opere dentro de un marco respetuoso de los derechos humanos, aunque ello reduzca la rentabilidad del modelo de negocio. Se trata de una metodología de diseño que debe aplicarse durante el desarrollo de las tecnologías de IA y no luego de que han sido implementadas (Penney *et al*., 2018). Esta política de prevención podría además incluir, entre otras obligaciones, la de ofrecer un "manual" para el uso de la nueva herramienta que visibilice las variables utilizadas para construir el modelo, señale los posibles sesgos en que podría incurrir y establezca

los protocolos de seguimiento y auditoría para facilitar la depuración y actualización del algoritmo con los mejores datos (Flórez, 2020).

Los algoritmos artificiales pueden mejorarnos la vida, pero también pueden empeorar nuestras sociedades cuando no responden a protocolos de transparencia en su diseño, no definen con claridad los objetivos que persiguen y no son objeto de un seguimiento constante que los evalúe y mejore.

Pensemos en que Waze podría incorporar variables de seguridad y sostenibilidad ambiental para evitar los cuadrantes más peligrosos y sugerir rutas más demoradas pero más costo-eficientes en términos ecológicos. Que Netflix podría permitir que el usuario active una función de "descubrimiento de nuevos géneros" cuando quiera innovar en sus gustos. Y que Rappi podría ofrecer una funcionalidad de "alimentación sana" para alertar al comensal cuando está consumiendo demasiada azúcar o calorías en exceso y sugerirle alternativas más saludables. Estos debates desde luego no son solo técnicos, sino morales, y por eso exigen una aproximación desde la ética y el derecho.

6. Conclusiones

El potencial de la biometría facial no se agota en su capacidad para disuadir al probable delincuente que se sabe vigilado, capturar al que ya cometió el delito, agilizar investigaciones forenses, controlar los procesos migratorios y encontrar a niños y adultos extraviados, tarea en la que ya demostró ser tremendamente eficiente en la India (Nagaraj, 2020 y NDTV, 2018).

En el sector de la salud, las TRF prometen servir para detectar la presencia de enfermedades y prevenir su avance en personas que aún no saben que las tienen, mediante el análisis de características del rostro (Lubin, 2016). Gracias al análisis facial computarizado ya es posible rastrear con mayor precisión el uso de medicamentos por parte de un paciente, apoyar los procedimientos de manejo del dolor y detectar enfermedades genéticas como el Síndrome de DiGeorge con una tasa de éxito del 96,6 % (Kruszka *et al.*, 2017).

En el sector bancario y de la venta al detal, la biometría facial se está posicionando como medio de autenticación en ban-

cos y cajeros automáticos, de apertura digital de cuentas y medio de pago tanto virtual como presencial sin necesidad de contacto físico en los establecimientos de comercio (Thales Group, 2021). El potencial de negocios es tan amplio, que se estima que el mercado global de las TRM habrá pasado de 3800 millones de dólares en 2020 a 8500 en 2025, con una tasa de crecimiento anual del 17,2 % (Mehra, 2020).

El potencial de explotación de las TRF para mejorar la vida de las personas es enorme. Sin embargo, como ocurre con la mayoría de tecnologías disruptivas, en sus etapas de despegue las acompañan múltiples riesgos que es prioritario identificar para lograr mitigar, bien sea mediante marcos éticos y jurídicos que regulen su funcionamiento o a través de mejoras técnicas.

En este trabajo exploramos específicamente los riesgos de la TRF cuando es implementada para acompañar la videovigilancia. De nuestro a análisis se desprenden principalmente cinco conclusiones:

1. Es posible identificar tres etapas de desarrollo técnico y regulatorio en la mayoría de nuevas tecnologías: la etapa de la *innovación*, la del *despegue* y la de *sofisticación* de la nueva tecnología (ver el detalle *at supra*, en la segunda parte). Con las TRF asistimos en la mayoría de los países a la etapa del despegue, que se caracteriza por una expansión acelerada del uso de la tecnología, que se ve acompañada de prácticas autorregulatorias (guías, procedimientos, auditorías y marcos éticos), de la conformación de redes de políticas públicas alrededor del funcionamiento la tecnología, pero también de la ausencia de legislación formal vinculante sobre la materia.

2. La reacción regulatoria frente a los retos que plantean las TRF ha sido asimétrica, según la región y el país del que se trate. Actualmente, se advierte una proliferación global de marcos éticos que buscan encuadrar la implementación de estas nuevas tecnologías, pero un déficit de regulación jurídica en la mayoría de los países. Los desarrollos éticos y las buenas prácticas autorregulatorias son una parte importante y deseable para construir un entorno responsable frente a la implementación de las TRF. Sin embargo, cuando no se complementan con legislación formal obligatoria pueden convertirse en un "distractor" que facilita la explotación de la tecnología, según los intereses empresariales y sin que el Estado las oriente hacia la consecución del bien común.

3. Se necesita con urgencia una agenda internacional general de protección de los derechos humanos frente a los riesgos que representa la inteligencia artificial y varias agendas específicas en función de cada especie de tecnología que pertenece al género de la IA. Las TRC son apenas una especie de esa familia de nuevas tecnologías y su uso plantea riesgos específicos de vulneración a los derechos humanos.

4. La implementación de las TRF genera riesgos de vulneración principalmente para tres derechos humanos: el derecho a la intimidad, materializado en el *habeas data* y la protección de los datos personales biométricos; el derecho a la no discriminación, que se ve amenazado por la posible implementación de sesgos algorítmicos injustos; y algunos derechos políticos, en particular el derecho a la protesta pacífica, el cual se puede ver afectado cuando, en contextos políticos autoritarios, las TRF son utilizadas por las agencias de seguridad del Estado para perfilar, estigmatizar, encerrar e incluso desaparecer a miembros de grupos sociales vulnerables, manifestantes y opositores de los gobiernos.

5. Entre los riesgos de vulneración de los derechos humanos detectados, la protección de datos admite una solución jurídica y de diseño de modelos seguros de gobernanza de datos que garanticen la confidencialidad de los datos biométricos. El sesgo algorítmico, por su parte, admite una solución tanto técnica como jurídica: técnica porque a medida que las TRF se desarrollan gracias al avance tecnológico, los sesgos injustos y los errores en la identificación automatizada tienden a cero; y jurídica porque solo la reglamentación de espacios de auditoría, supervisión y evaluación constante de las TRF por parte de expertos y de la comunidad en general puede garantizar su mejora continua e implementación transparente. Por último, la dinámica de la amenaza al derecho a la protesta responde al contexto político en que se desarrolla: en los Estados autoritarios es imposible proteger este derecho frente a los abusos del gobierno; en los Estados democráticos, el reto consiste en contener la discrecionalidad de la fuerza pública y establecer protocolos de seguridad que prevengan el mal uso de los datos biométricos.

En definitiva, las TRF y los SVRF se revelan como un *carrefour* de convergencia tanto de riesgos como de oportunidades para la vulneración y la protección de los derechos humanos. Por ello, desde el momento de la concepción de cualquier sistema de videovigilancia se debe buscar una protección integral de los derechos humanos mediante la adopción de un enfoque

de *human-rights-by-design* (Flórez, 2020), así como procurar la consideración de todas las demás aspiraciones sociales legítimas que rodean a un proceso de construcción de política pública que implica disrupción tecnológica.

Una sociedad hipervigilada por tecnologías de inteligencia artificial supone un desafío mayúsculo para la ética digital y el derecho. Por consiguiente, diversos sectores de la sociedad deben enfrentar articuladamente el reto regulatorio que nos plantea la disrupción tecnológica de la IA. Esta investigación fue una oportunidad para hacerlo desde la academia y para contribuir así a enriquecer la conversación que hoy avanza a toda velocidad en América Latina y el mundo sobre las buenas prácticas y salvaguardas éticas, jurídicas y técnicas necesarias para utilizar la tecnología como un vehículo del cambio social positivo sujeto a controles, pero sin frenar la innovación.

7. Referencias

Allison-Hope, D. (2017). Human Rights by Design. *BSR*. https://bit.ly/3FxO4hY

Amnistía Internacional (2020). *Rusia: No se debe usar tecnología invasiva de reconocimiento facial para reprimir protestas*. https://bit.ly/3j53YqI

Angwin, J., Larson, J., Mattu, S. y Kirchner, L. (2016). Machine bias there's software used across the country to predict future criminals. and it's biased against blacks. *Propublica*. https://bit.ly/3jqAnsb

Armitage, R. (2002). To CCTV of Not to CCTV: a review of current research into the effectiveness of CCTV systems in reducing crime. *Nacro Community Safety Practice Briefing*, 1-8. https://bit.ly/3suArKq

Barbosa, D., Fetzer, T., CL Souza, P. y Vieira, C. (2021). *De-escalation technology: the impact of body-worn cameras on citizen-police interactions*. London, Centre for Economic Policy Research. https://bit.ly/3mgnK31

Becker, G. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76, 2, 169-217.

Bischoff, P. (2021). Surveillance camera statistics: which cities have the most CCTV cameras? *Comparitech*. https://bityl.co/8MR3

Botsman, R. (2017). Big data meets Big Brother as China moves to rate its citizens. *Wired*. https://bityl.co/8P7s

Buolamwini, J. y Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, 81, 1-15. https://bityl.co/8PrQ

Castaño, D. (2020). "La gobernanza de la inteligencia artificial en América Latina: entre la regulación estatal, la privacidad y la ética digital", en *Inteligencia Artificial en América Latina y el Caribe. Ética, gobernanza y políticas* (pp. 5-33), CETyS.

Castelvecchi, D. (2020). Beating biometric bias. *Nature*, (587), 347-349. https://bityl.co/7flY

Chanthadavong, A. (2020). New Zealand establishes algorithm charter for government agencies. *ZDNet*. https://zd.net/3mMkpKP

Crumpler, W. (2020). How Accurate are Facial Recognition Systems – and Why Does It Matter? *Center for Strategic and International Studies*. https://bityl.co/8P8K

Collins, J. (2017). Cathy O'Neil's weapons of math destruction: how big data increases inequality and threatens democracy. *Jason Collings Blog*. https://bit.ly/3Dxf0NW

Cortés, N. (2018). Cuatro 'rutas de la muerte' que fueron sugeridas por el celular. *El Tiempo*. https://bit.ly/3jsLkcN

Dammert, L. y Silva, A. (2018). *Seguridad y tecnología en América Latina: experiencias y desafíos*. Universidad de Santiago de Chile.

Donahoe, E. y MacDuffee, M. (2019). Artificial intelligence and human rights. *Journal of Democracy*, 30(2), 115-126. https://doi.org/ghv2rv

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://reut.rs/3yx84wm

Doffman, Z. (2019). Hong Kong Exposes Both Sides of China's Relentless Facial Recognition Machine. *Forbes*. https://bityl.co/8MRF

Duranton, S., Erlebach, J., Brégé, C., Danziger, J., Gallego, A. y Pauly, M. (2020). What's keeping women out of data science? *BCG*. https://on.bcg.com/3gleJxC

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Eubanks, V. (2019). La Automatización de los Perjuicios. *Investigación y Ciencia*. https://bit.ly/3iZPE2M

Farrenas, C. (2021). Protección de datos rechaza el uso del reconocimiento facial en exámenes online. *La Vanguardia*. https://bit.ly/3DlRdcY

Feldman, L. Adolphs, R., Marsella, S., Martinez, A. y Pollak, S. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, *20*(1), 1-68.

Feldstein, S. (2019). *The Global Expansion of AI Surveillance*. Carnegie Endowment for International Peace.

Flórez, J. F. (2020). Inteligencia artificial y derechos humanos. *Revista Semana*. https://bit.ly/3gcJb2T

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. y Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines, 28*, 689-707. https://bit.ly/399QVyy

Fussell, S. (2018). *North Korea may be secretly selling face recognition tech abroad through front companies*. https://bit.ly/3z6YjpI

Gates, K. (2011). *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*. New York University Press.

Gómez, S., Mejía, D. y Tobón, S. (2017). *The deterrent effect of public surveillance cameras on crime.* Universidad EAFIT.

Guío Español, A. (2021). *Modelo de gobernanza de la infraestructura de datos para el desarrollo de tecnologías emergentes*. Gobierno de Colombia. https://bit.ly/3z4TgGx

Guío Español, A., Tamayo Uribe, E., Gómez Ayarbe, P. y Mujica, M. P. (2021). *Marco ético para la inteligencia artificial en Colombia*. https://bit.ly/37sRcPx

Harari, Y. N. (2020). Yuval Noah Harari: the world after coronavirus. *Financial Times*. https://on.ft.com/3AVt29N

Hesseling, R. (1994). Displacement: A review of the empirical literature. *Crime Prevention Studies*, 3, 197-230.

Interpol (2020). Reconocimiento facial. https://bityl.co/8Pi8

Jasso L. y Jasso, C. (2021). Abuso policial, discrecionalidad y tecnologías de vigilancia en América Latina. *Iztapalapa Revista de Ciencias Sociales y Humanidades, 42*(90), 119-144.

Jasso, L. (2020). Seguridad ciudadana y tecnología: uso, planeación y regulación de la videovigilancia en Latinoamérica. *Revista de Investigación en Derecho, Criminología y Consultoría Jurídica, 14*(27), 5-27.

Kosinski, M. (2021). Facial recognition technology can expose political orientation from naturalistic facial images. *Nature, Sci Rep, 11,* 100. DOI:10.1038/s41598-020-79310-1.

Kruszka, P., Addissie, Y., McGinn, D. E., Porras, A. R., Biggs, E., Share, M., Blaine, C. Chung, B., Mok, G., Mak, C, Muthukumarasamy, P., Thong, M.-K. Sirisena, N. D., Dissanayake, V. Paththinige, S., Prabodha, L. B. Mishra, R,. Shotelersuk, V., Nsikaka, E., [...], Muenke, M. (2017). 22q11.2 deletion syndrome in diverse populations. *American Journal of Medical Genetics*, *173*(4), 879-888. https://doi.org/f93x53

Lara, J. C. (2020). "Perspectivas de derechos humanos en la implementación de marcos éticos para la inteligencia artificial". En C. Aguerre (ed.). *Inteligencia Artificial en América Latina y el Caribe. Ética, gobernanza y políticas* (pp. 34-66). CETyS y UdeSA.

Lio, V. (2015). Ciudades, cámaras de seguridad y video-vigilancia: estado del arte y perspectivas de investigación. *Astrolabio*, *15*, 273-302.

Lubin, G. (2016). Facial profiling could be dangerously inaccurate and biased, experts warn. *Business Insider*. https://bit.ly/2W0zNZn

McLaughlin, M. y Castro, D. (2020). The critics were wrong: NIST data shows the best facial recognition algorithms are neither racist nor sexist. *Information Technology and Innovation Foundation*. https://bit.ly/3kB34lu

Mehra, A. (2020). Facial recognition market worth $8.5 billion by 2025. *Markets and Markets*. https://bit.ly/2YbRE03

Mendoza, S. (2019). San Francisco, primera ciudad en prohibir la tecnología de reconocimiento facial en EE. UU. *El País*. https://bit.ly/37YrYWg

Miailhe, N. (2018). The geopolitics of artificial intelligence: The return of empires? *Politique Étrangère,* (3), 105-117.

Mozur, P. (2018). Inside China's dystopian dreams: A.I., shame and lots of cameras. *The New York Times*. https://nyti.ms/3D4Gk5F

Mozur, P. (2019a). One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority. *The New York Times*. https://bityl.co/8MRe

Mozur, P. (2019b). In Hong Kong Protests, Faces Become Weapons. *The New York Times*. https://bityl.co/8MRo

Nagaraj, A. (2020). Indian police use facial recognition app to reunite families with lost children. *Reuters*. https://reut.rs/3B1FKEh

Nájera, J. y Ricaurte, P. (2021). Tecnologías de interés público: el caso de las coronapps en América Latina. *Centro Latam Digital*. https://bit.ly/3sybQo7

NDTV (2018). Facial Recognition System Helps Trace 3,000 Missing Children In 4 Days. https://bityl.co/8P88

Norris, C., McCahill, M. y Wood, D. (2004). The Growth of CCTV: a global perspective on the international diffusion of video surveillance in publicly accessible space. *Surveillance & Society*, 2(2/3), 110-135.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy,* Penguin.

Ortega, A. (2021). Hacia un régimen europeo de control de la Inteligencia Artificial. *Real Instituto Elcano.* https://bit.ly/3keO5h1

Pascual, M. (2021). Cuando el algoritmo se equivoca. *El País.* https://bit.ly/3y4DAlf

Penney, J., McKune, S., Gill, L. y Deibert, R. J. (2018). Advancing Human-Rights-By-Design In The Dual-Use Technology Industry. *Journal of International Affairs.* https://bit.ly/3uq0Olu

Prakash, A. (2018). *Go.AI (Geopolitics of Artificial Intelligence).* Abishur Prakash.

Raji, I., Gebru, T., Mitchell, M., Boulamwini, J., Lee, J. y Denton, E. (2020). Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES '20),* arXiv:2001.00964, 145-151. https://doi.org/gg4hpd

Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C. y Kim, L. (2018). *Artificial intelligence & human rights: opportunities and risks.* Harvard Law School.

Ratcliffe, J. Taniguchi, T. y Taylor, R. (2009). The crime reduction effects of public CCTV cameras: A multi-method spatial approach. *Justice Quarterly, 26*(4), 746-770.

Raz, D., Bintz, C., Guetler, V., Tam, A., Katell, M., Dailey, D., Herman, B., Krafft, P. M., Young, M (2021). Face Mis-ID: An Interactive Pedagogical Tool Demonstrating Disparate Accuracy Rates in Facial Recognition. *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 845-904. https://shortdoi.org/

Reevell, P. (2020). How Russia is using facial recognition to police its coronavirus lockdown. *ABC News.* https://abcn.ws/3y1lUqO

Risse, M. (2018). Human rights and artificial intelligence: an urgently needed agenda. *HKS Faculty Research Working Paper Series.* https://bit.ly/2UXeKWU

Risse, M. (2019). Human Rights and Artificial Intelligence: An Urgently Needed Agenda. *Human Rights Quarterly, 41*(1), 1-16. https://doi.org/gg88vf

Roussi, A. (2020). Resisting the Rise of Facial Recognition. *Nature*, 587, 350-353. https://go.nature.com/3k5QryM

Smith, A. (2019). More Than Half of U.S. Adults Trust Law Enforcement to Use Facial Recognition Responsibly. *Pew Research Center.* https://pewrsr.ch/3igPD9X

The Physics arXiv Blog (2014). The Face Recognition Algorithm That Finally Outperforms Humans. https://bit.ly/2Wvv1mO

The Technolawgist (2020). Nueva Zelanda, un ejemplo de transparencia y responsabilidad en el uso de los datos por parte del sector público. https://bityl.co/8P9H

Velasco, L. (2019). Gobernar la inteligencia artificial: el reconocimiento facial, *Agenda Pública*. https://bit.ly/3zbeany

Wee S. y Chen, E. (2019). China's tech firms are mapping pig faces. *The New York Times*. https://nyti.ms/3z2kvkN

Wagner, B. (2018). Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications, Consejo de Europa, Comité de Expertos en Intermediarios de Internet (MSI-NET).

Webb, A. (2020). *Nueve gigantes. Las máquinas inteligentes y su impacto en el rumbo de la humanidad*. Paidós.

Webster, W. (2004). The diffusion, regulation and governance of closed-circuit television in the UK. *Surveillance & Society, 2*(2/3), 230-250.

Welsh, B. y Farrington, D. P. (2009). Public area CCTV and crime prevention: An updated systematic review and meta-analysis. *Justice Quarterly*, 26(4), 716-745.

Wiggers, K. (2021). Outlandish Stanford Facial Recognition Study Claims There Are Links Between Facial Features and Political Orientation. *Venturebeat*. https://bit.ly/3z1EZKM

Williams, G. (2019). Videovigilancia en espacios públicos. *Biblioteca del Congreso Nacional de Chile*. https://bit.ly/2Wj84mU

Referencias normativas y judiciales

Comisión Europea (2021a). *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Fostering a European approach to artificial intelligence*. https://bit.ly/3ghgUbq

Comisión Europea (2021b). *Proposal for a regulation of the European Parliament and of the Council, laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts.* https://bit.ly/2UzTrut

Consejo Nacional de Política Económica y Social (2018). *Documento CONPES 3920. Política Nacional de Explotación de datos (Big Data).* https://bit.ly/2W3nziH

Consejo Nacional de Política Económica y Social (2019). *Documento CONPES 3975. Política Nacional para la transformación digital e inteligencia artificial*. https://bit.ly/38eKWIB

Corte Constitucional de la República de Colombia. *Sentencia T-407/12* (31 de mayo de 2021) [M. P. Mauricio González Cuervo].

European Data Protection Board y European Data Protection Supervisor (2021). *EDPB & EDPS call for ban on use of AI for automated recognition of human features in publicly accessible spaces, and some other uses of AI that can lead to unfair discrimination.* https://bit.ly/3B0WtXY

Naciones Unidas (2011). *Principios Rectores sobre las Empresas y los Derechos Humanos*, Nueva York y Ginebra.

Parlamento Europeo (2016). *Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo.* https://bit.ly/3ie9GWJ

Parlamento Europeo (2018a). *Reglamento (UE) 2018/1860 del Parlamento Europeo y del Consejo.* https://bit.ly/3IWGvZe

Parlamento Europeo (2018b). *Reglamento (UE) 2018/1861 del Parlamento Europeo y del Consejo*. https://bit.ly/3COVyuP

Parlamento Europeo (2018c). *Reglamento (UE) 2018/1862 del Parlamento Europeo y del Consejo*. https://bit.ly/2ZAZUaG

Sala Novena de Revisión de la Corte Constitucional (31 de julio de 2008). *Sentencia T-768/08* [M. P. Clara Inés Vargas Hernández].

Sala Plena de la Corte Constitucional de la República de Colombia (3 de marzo de 2020). *Sentencia C-094/20* [M. P. Alejandro Linares Cantillo].

Sala Primera de Revisión de la Corte Constitucional de la República de Colombia (3 de abril de 2018). *Sentencia T-114/18*. [M. P. Carlos Bernal Pulido].

Sala Séptima de Revisión de Tutelas de la Corte Constitucional (28 de julio de 2017). *Sentencia T-487/17* [M. P. Alberto Rojas Ríos].

Superintendencia de Industria y Comercio (s. f.). *Protección de datos personales en sistemas de videovigilancia*. https://bit.ly/3mjygZ1

Informes y reportes

Asociación por los Derechos Civiles (2017). *Cuantificando identidades en América Latina. Un breve repaso acerca de cómo las sociedades latinoamericanas se enfrentan a la implementación de las tecnologías biométricas*.

Banco Interamericano de Desarrollo y C Minds (2020). *La inteligencia artificial al servicio del bien social en América Latina y el Caribe*. https://bit.ly/3og44Py

Grother, P., Ngan, M. y Hanaoka, K. (2018). *Ongoing Face Recognition Vendor Test (FRVT). Part 2: Identification. National Institute of Standards of Technology. U. S. Department of Commerce*. https://doi.org/g2mm

Grother, P., Ngan, M. y Hanaoka, K. (2021). Ongoing Face Recognition Vendor Test (FRVT). *Part 6B: Face recognition accurancy with face masks using post-COVID-19 algorithms. National Institute of Standards of Technology*. https://bit.ly/3gMVALg

Thales Group (2021). *Facial recognition: top 7 trends (tech, vendors, markets, use cases & latest news)*. https://bit.ly/3gFR29s

The Economist Intelligence Unit (2016). Artificial Intelligence in the Real World. The Business Case Takes Shape. *Inside Bigdata*. https://bit.ly/3sWnclN

Human-Centered Artificial Intelligence Institut of Stanford University (2021). *Artificial intelligence index report 2021*. Stanford University.

Human-Centered Artificial Intelligence Institut of Stanford University (2019). *Artificial intelligence index: 2019 annual report*. Stanford University.

Sitios web

https://carnegieendowment.org/publications/interactive/ai-surveillance

https://www.fightforthefuture.org

http://gendershades.org

https://reconocimientofacial.info

https://www.statewatch.org

Descargo de responsabilidad. *Las opiniones expresadas en la publicación incumben únicamente a los/as autores/as. No tienen intención de reflejar las opiniones o perspectivas del CETyS, CLD ni de ninguna otra organización involucrada en el proyecto.*