RIESGOS JURÍDICAMENTE RELEVANTES E INTELIGENCIA ARTIFICIAL: HACIA UNA REGULACIÓN MÁS HUMANA EN UN MUNDO CADA VEZ MÁS ARTIFICIAL

Presentado por:

Juan José Romero Morales

Trabajo de grado para optar por el título de abogado

Tutor:

Majer Nayi Abushihab Collazos

COLEGIO MAYOR DE NUESTRA SEÑORA DEL ROSARIO - UNIVERSIDAD DEL ROSARIO

FACULTAD DE JURISPRUDENCIA

BOGOTÁ, 2023

Contenido

Tabla de Abreviaturas:

ABREVIATURA	SIGNIFICADO
CA	Cambridge Analytica
CAS	Escándalo de Cambridge Analytica
DF	Deepfake
DL	Deep Learning
EC	Comisión Europea
ECOA	The Equal Credit Opportunity Act
EE. UU.	Estados Unidos de América
FBI	Oficina Federal de Investigacio
IA (AI)	Inteligencia Artificial
IIAA	Inteligencias Artificiales
IOGANA	Identifying Outputs of Generative Adversarial Networks Act
ML	Machine Learning
NDAA	U.S. National Defense Authorization Act
NLP	Natural Language Processing
NYDFS	Departamento de Servicios Financieros de Nueva York
OECD	Organización para la Cooperación y el Desarrollo Económicos
ONG	Organismo No Gubernamental
RU	Reino Unido
UE	Unión Europea
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

Resumen.

En esta monografía, a través de la revisión de casos, se analizan los riesgos jurídicos asociados al uso y desarrollo de Inteligencias Artificiales. A partir de este análisis, se proponen una serie de principios rectores para el desarrollo de modelos normativos destinados a la administración de los riesgos identificados. Los resultados de la investigación muestran la **existencia** de importantes riesgos jurídicos en áreas como la administración de justicia, la seguridad nacional y la seguridad social, y la necesidad de un marco normativo global para abordarlos. En conclusión, la necesidad de un sistema normativo basado en los principios de: desarrollo *pro persona*, especialidad de materia, precaución, actualización y transparencia material, será necesario para administrar y mitigar los riesgos asociados con las IIAA.

Abstract.

In this monograph, through the review of cases, the legal risks associated with the use and development of Artificial Intelligences are analyzed. From this analysis, a series of guiding principles are proposed for the development of regulatory models to manage the identified risks. The research results show the existence of significant legal risks in areas such as the administration of justice, national security and social security, and the need for a comprehensive regulatory framework to address them. In conclusion, the need for a regulatory system based on the principles of: pro persona development, subject matter specificity, precaution, updating and material transparency, will be necessary to manage and mitigate the risks associated with IIAAs.

1. Introducción

> *"La inteligencia artificial puede hacer posible lo que antes parecía imposible, pero también puede plantear nuevas amenazas. La regulación es necesaria para minimizar el riesgo y maximizar el potencial de la IA."* (ChatGPT)

Es interesante analizar cómo, para sociólogos de la talla de Beck (1993), el debate público frente a la gestión de los riesgos, introducidos en nuestro día a día por el avance tecnológico, es un elemento central de la "nueva modernidad" en la que nos encontramos. De acuerdo con el autor, esto permite un tránsito hacía una modernidad mucho más responsable, toda vez que la carga de responsabilidad se ubica en el productor del riesgo y no en la victima del mismo. A partir de lo esto es posible apreciar la transición de un sistema de riesgos controlados de manera privada, y conveniente, a uno mucho más democrático. Finalmente, indicaría que, en consecuencia, la construcción de un sistema de normas jurídicas y valores sociales se ve amplificado, lo que genera la posibilidad la legitimación y estabilización de los sectores, públicos o privados, implicados tanto en la producción de las fuentes de riesgo como en la gestión de estas.

Si bien existen diferencias entre las posiciones de los variados autores que han analizado esta "modernidad reflexiva", respecto a quienes deben representar la institucionalidad; es evidente, que concuerdan en la necesidad en la construcción de una confianza individuo-institución generada, principalmente, a partir de un proceso dialógico entre las partes, como lo expresa Giddens (1998). Frente a esta última posición, para el autor británico, la existencia de una sociedad activa, propioceptiva y comprometida con su subsistencia es indispensable.

Esto cobra mayor importancia cuando consideramos factores como los abordados por Cohen & Méndez (2000) en su obra. Es así como, para las autoras, la globalización, el contexto de riesgo permanente y la incertidumbre, que caracterizan la situación social contemporánea, son factores que generan el nacimiento de movimientos sociales típicos de las sociedades reflexivas, los cuales asumen un papel protagónico en la ruptura de estas barreras e impulsan cambios en el statu quo.

Es en esa medida que el impulso de espacios de dialogo entre la ciudadanía y las instituciones es fundamental para la construcción de un orden social mucho más seguro para todos sus integrantes. Ahora bien, la identificación de nuevos riesgos, así como la carga de responsabilidad que estos acarrean, es una condición sin la cual no podría darse un inicio al ejercicio dialógico entre las diferentes partes que componen a la sociedad. Por lo tanto, dicha tarea debe ser asumida por los sectores interesados en la generación de cambios. A partir de lo expuesto, y teniendo en cuenta los objetivos de democratización que se plantean autores como Beck (1993) y Giddens (1998), se evidencia que el papel de sectores académicos pertenecientes a las ciencias sociales será fundamental a la hora de humanizar los procesos de desarrollo e implementación de nuevas tecnologías, lo anterior con el fin de mitigar los riesgos que podrían ser introducidos a la sociedad.

Con estos preceptos como base, se evidencia, por lo menos preliminarmente, la necesidad social de identificar la relevancia de las diferentes fuentes de riesgo, actuales o inminentes, presentes en el que hacer social, para así proponer planes de contingencia y/o acción para un manejo seguro de estas. Como se expondrá más adelante, una de las industrias con mayor crecimiento en la actualidad y con mayor aplicación es la de tecnologías de la automatización; sin embargo, debido a la velocidad a la que se desarrolla, es uno de los sectores tecnológicos

con menor regulación. Tan solo hasta el 2021 la Unión Europa (UE) presentó una propuesta para el desarrollo de un régimen legal relacionado con las inteligencias artificiales (IIAA). De acuerdo con lo observado por la Comisión Europea[1] (2021a), existen diferentes problemáticas que se derivan del desarrollo e implementación de las inteligencias artificiales. Dichas situaciones podrían ser agrupadas en dos grupos: (a) situaciones de riesgos personales y (b) situaciones de riesgos económicos. Entre los mencionados supuestos de hecho cabe resaltar la creación y la intensificación de riesgos relativos al impacto en derechos fundamentales de las personas expuestas al funcionamiento de las IIAA y por otra parte la incertidumbre relacionada a la atribución de responsabilidades que afrontan las empresas, lo cual implica falta de confianza y por lo tanto menor desarrollo del sector económico.

Ahora bien, como se puede observar la EC (2021b) ha identificado una jerarquía de riesgos relacionados con la funcionalidad de las IIAA, indicando que la intervención estatal directa en los programas de "riesgo alto" estaría justificada. Vale la pena resaltar que los sistemas de alto riesgo son aquellos que automatizan procesos relacionados con derechos como: salud, educación, acceso al trabajo, justicia, acceso al sistema financiero, entre otros. No obstante, dicho marco jurídico sería adoptado a finales del 2023, y únicamente de manera transitoria. Según proyecciones de la EC (2021c) solamente se podría hablar de una aplicabilidad real del marco jurídico hasta la segunda mitad de 2024.

Esto demuestra la relevancia y actualidad del debate frente a la regulación del uso y desarrollo de IIAA; así como la inevitabilidad de la regulación a futuro. No obstante, y tal como lo señalan Candelon et al (2021), aún es necesario mucho trabajo para plantear un

[1] EC por sus siglas en inglés.

modelo de intervención eficiente que permita mitigar los posibles daños derivados de actividades que impliquen a las IIAA, sin ser un lastre para la industria. Es por este motivo, que la presente monografía pretende responder la siguiente pregunta ¿Cuáles deberían ser los principios rectores generales de un modelo de intervención y mitigación del riesgo sobre las inteligencias artificiales? Para cumplir con este fin: (a) se determinará el concepto de riesgo relevante; (b) se analizarán diferentes escenarios en los que las IIAA están siendo empleadas actualmente; (c) se establecerá si los escenarios analizados corresponden a situaciones de riesgo; (d) se analizará la posición que algunos estados han adoptado en la actualidad; y (e) se plantearán observaciones relativas a las puntos analizados y se presentará una propuesta de principios que deberían regir las futuras regulaciones relacionadas con la IIAA.

2. La amplificación de un riesgo

2.1. Sobre los riesgos en el estado actual de la sociedad

Determinar la legitimación de las intervenciones del Derecho en la vida social constituye uno de los elementos de mayor análisis, y a su vez importancia, para la teoría jurídica. En esa medida, tesis como la argumentada por Rousseau (1762), entre otros contractualistas, han determinado como la sociedad ha renunciado a una serie de libertades, para conformar un ente capaz de administrar y resolver los conflictos que se originan desde la tensión entre los participantes de este contrato social. A partir de estas ideas, doctrinantes de diferentes áreas del Derecho han centrado sus análisis en los mecanismos de los que debería valerse el Estado para solucionar los enunciados conflictos. Es así, que la pregunta frente a ¿cuál es la manera

más justa para imputar responsabilidad a una persona? Ha sido desarrollado en una gran proporción.

Actualmente, es claro que la tesis de la imputación objetiva ha sido acogida en gran medida como el mecanismo de atribución de responsabilidad que permite mayor eficiencia al momento de determinar la causalidad entre la conducta de una persona y un eventual resultado lesivo, como lo ha expuesto Jakobs (1996). Ahora bien, en su obra, el autor alemán desarrolla dos elementos que a su juicio requieren de mayor grado de claridad: (a) el riesgo jurídicamente permitido; y (b) la prohibición de regreso (pág. 10). Desde la perspectiva del autor, "una sociedad saturada por la técnica esperará de un fabricante de máquinas que éste no cree nuevos riesgos, y por tanto le impondrá el deber de garantizar la inocuidad en todas las condiciones de funcionamiento" (Jakobs, 1996: 17) mientras que "una sociedad que esté necesitada de avances técnicos tolerará bastantes riesgos" (*Ibidem)*. Es en esa medida, que será necesario determinar sociológicamente en qué estado se encuentra una comunidad puntual, con el fin de atribuir, satisfactoriamente, la eventual responsabilidad causal entre la conducta de uno de sus miembros y el resultado de esta.

Es a partir de lo anteriormente planteado, que cobra importancia el análisis desarrollado por Ulrich Beck en la década de los 80's. De acuerdo con Beck:

> De una manera similar a como en el siglo XIX la modernización disolvió la sociedad agraria anquilosada estamentalmente y elaboró la imagen estructural de la sociedad industrial, la modernización disuelve hoy los contornos de la sociedad industrial, y en la continuidad de la modernidad surge otra figura social. (Beck, 1986: 16)

Es de esta manera, que el autor introduce la noción de La sociedad del riesgo. En esta medida, la sociedad genera una realidad en que los efectos secundarios latentes de sus

actividades se vuelven tan perceptibles que inicia un debate sobre la necesidad de gestionar dichos efectos, sin obstaculizar el avance técnico que estos presentan, pero tampoco afectar en sobremedida a la población. Así mismo, para Beck (1986) esto implica la necesidad de generar nuevas medidas de seguridad, demandadas por una comunidad alerta.

No obstante, es evidente que el análisis desarrollado por Beck (1986) se basa en un tipo de riesgo de naturaleza sistémica y generalizada; es por esto, que el autor caracteriza esta nueva sociedad como dentro de la superación del riesgo personal y el nacimiento del riesgo global. En esta medida, el autor propone cinco tesis que desarrolla a lo largo de su escrito de las cuales vale la pena resaltar:

1. Los riesgos que se generan en el nivel más avanzado del desarrollo de las fuerzas productivas (…) se diferencian esencialmente de las riquezas. Estos riesgos causan daños sistemáticos y a menudo irreversibles, suelen permanecer invisibles, se basan en interpretaciones causales, por lo que sólo se establecen en el saber (científico o anticientífico) de ellos, y en el saber pueden ser transformados, ampliados o reducidos, dramatizados o minimizados, por lo que están abiertos en una medida especial a los procesos sociales de definición. Con ellos, los medios y las posiciones de la definición del riesgo se convierten en posiciones sociopolíticas clave.
2. Con el reparto y el incremento de los riesgos surgen situaciones sociales de peligro. Ciertamente, en algunas dimensiones éstas siguen a la desigualdad de las situaciones de clases y de capas, pero hacen valer una lógica de reparto esencialmente diferente: los riesgos de la modernización afectan más tarde o más

temprano también a quienes los producen o se benefician de ellos. Contienen un efecto bumerang que hace saltar por los aires el esquema de clases. Tampoco los ricos y poderosos están seguros ante ellos. Y esto no sólo en tanto que peligros para la salud, sino también en tanto que peligros para la legitimación, la propiedad y la ganancia (…). Al mismo tiempo, los riesgos producen nuevas desigualdades (…)

5. (…) lo que hasta el momento se había considerado apolítico se vuelve político: la supresión de las «causas» en el proceso de industrialización mismo. De repente, la opinión pública y la política empiezan a mandar en el ámbito íntimo del management empresarial, en la planificación de la producción, en el equipamiento técnico, etc. Ahí queda claro de una manera ejemplar de qué se trata propiamente en la disputa pública sobre la definición de los riesgos: no sólo de las consecuencias para la salud de la naturaleza y del ser humano, sino de los *efectos secundarios sociales, económicos y políticos de estos efectos secundarios:* hundimiento de mercados, desvalorización del capital, controles burocráticos de las decisiones empresariales, apertura de nuevos mercados, costes monstruosos, procedimientos judiciales. En la sociedad del riesgo surge así a impulsos pequeños y grandes (…) el *potencial político de las catástrofes.* La defensa y administración de las mismas puede incluir una *reorganización del poder y de la competencia.* (Beck, 1986: 28-30)

A partir de las tesis expuestas, queda en evidencia que el interés por la gestión y la mitigación de los riesgos se convierte en uno de los principales intereses de cada una de las partes que conforman el mercado. Sobre todo, cuando los riesgos introducidos por el desarrollo industrial moderno son tan intensos y generalizados que sus impactos son

perceptibles, a nivel individual, por casi que cualquier persona, como lo señalan Cohen & Méndez (2000).

Si tomamos el estado de la sociedad propuesto por Beck junto a la consideración desarrollada por Jakobs, se encontrará que el valor, o más bien el desvalor, que le damos a los riesgos actualmente responde a una lógica de permisividad limitada. Esto quiere decir que reconocemos la necesidad de coexistir con una serie de riesgos latentes y constantes, pero los controlamos para evitar su desbordamiento. En esa medida, el segundo elemento trabajado por el penalista en su tratado (la prohibición de regreso) cobra importancia. Jakobs (1996) define la prohibición de regreso como una fórmula para limitar la responsabilidad de las personas que, si bien, hicieron parte del curso causal que genero un resultado no les es atribuible el resultado de este (pág. 83). A partir de esto, es claro que dicha figura pretende identificar al o a los posibles responsables que, con su actuar, desbordaron en desmedida el riesgo socialmente permitido y por lo tanto generaron un resultado lesivo injusto.[2]

Con base en lo planteado, es evidente que al realizar dicho estudio de causalidad se logrará discriminar entre los diferentes casos en que el daño se generó a partir del actuar de los productores del riesgo o, por otra parte, si se debió a una conducta desplegada por los consumidores de este, logrando así ubicar correctamente la responsabilidad del fallo en un servicio y, en consecuencia, implementar las medidas correctivas más eficientes. Un ejemplo clásico, pero reducido, de dicho fenómeno puede encontrarse en los accidentes automovilísticos; en la situación hipotética en que un conductor atropelle a dos personas podría deberse a dos causas, la primera donde el sujeto al volante, en estado de embriaguez,

[2] En este punto resultado injusto debe entenderse como el resultado típico y antijurídico contemplado en la ley penal.

no logra frenar y termina con la vida de dos personas, o un segundo caso en el que el conductor, sobrio, reconoce a los peatones e intenta evitar el colapso, sin embargo los frenos del auto no responden, por una falla en su diseño, y se termina por concretar el resultado. En ambos casos tenemos el mismo resultado, sin embargo, es evidente que la responsabilidad se encuentra en cabeza de dos personas completamente distintas. Situaciones similares se presentan, en gran cantidad, en distintas áreas y bajo diferentes lógicas.

Habiendo determinado que el desbordamiento de los riesgos permitidos puede darse en uno u otro de los extremos de la cadena de consumo, es necesario preguntarse por las estrategias aplicables que cada persona, dentro de su rol, puede aplicar. Ahora bien, es importante hacer hincapié en que los riesgos "individuales" expuestos siguen siendo introducidos por el avance técnico-social actual y en esa medida deben seguir siendo gestionados paralelamente a los riesgos globales considerados por Beck.

2.2. La identificación de un riesgo ya presente

La identificación de riesgos socialmente relevantes representa uno de los principales motores para el desarrollo normativo de una comunidad. Esto se evidencia al estudiar fenómenos y/o conceptos como las actividades peligrosas, los sectores económicos regulados y la expedición de licencias especiales para poseer y operar ciertos bienes, entre otros. Es así como la sociedad, y en especial el legislador, se encuentra en la tarea constante de identificar los riesgos generados por cada actividad, con el fin de determinar la necesidad de intervención frente a la misma. Desde una perspectiva sociológica el arte representa una fuente de consulta importante. Dentro de las diferentes teorías estéticas, resaltan los postulados desarrollados por Nietzsche (1872), los cuales identifican al arte como una forma de expresión pasional con un resultado catártico, a través del cual, tanto el público como los

artistas son capaces de experimentar situaciones y procesar sentimientos dentro de un espacio contralado.

En gran medida, las representaciones artísticas nacen a partir de las preocupaciones de sus creadores. No obstante, estas funcionan como una forma de expresión colectiva de miedos e inseguridades generales a una comunidad, esto se debe a que socialmente generamos una narrativa compartida guida por códigos de lectura a través de los que procesamos la realidad, tal como lo expone Foucault (1966), es así como el autor identifica que, comunitariamente, logramos generar categorías que identificamos como lo propio o lo otro.

Ahora bien, esta monografía se ocupará de una fuente de riesgos socialmente identificada desde hace ya un tiempo. Obras notables que recogen las principales problemáticas que como sociedad hemos identifica frente a este sector son: Minority Report (1956),[3] I Robot (1950),[4] Terminator (1984), Blade Runner (1982) o Matrix (1999), entre muchas otras. Existen dos elementos característicos/comunes de estas obras. En primer lugar, todas se ambientan en un futuro distópico y desalentador para la humanidad y, en segundo lugar, la causa de esta distopia se debe a un desarrollo desenfrenado tecnológico que ha generado una nueva especie de máquinas capaces de antagonizar a los humanos. Si bien nos encontramos lejos del argumento central de algunas de estas obras, hay otros que actualmente estamos viviendo. Lo analizado genera la posibilidad de desarrollar un diagnóstico social, toda vez que permite identificar un miedo común: la tecnología y su avance.

No obstante, es evidente que muchos de estos miedos están basados en visiones apocalípticas del mundo. Sin embargo, el avance tecnológico actual trae consigo riesgos que

[3] Adaptada al cine en 2002.
[4] Adaptada al cine en 2004.

vale la pena explorar. En la actualidad, la aplicación de IIAA en distintos campos del conocimiento se ha convertido en una realidad. Sin embargo, no existe un consenso científico general frente a lo que son las IIAA, tal como lo ha expresado el *National Science and Technology Council Report* (2016). De acuerdo con Simon (2019), técnicamente, las IIAA están compuestas por un diverso número de tecnologías en las que sobresalen: *Machine Learning (ML), Deep Learning (DL), Computer Vision, Natural Language Processing (NLP), y Machine Reasoning*.

De las anteriores herramientas es importante hacer hincapié en el ML y el DP, esto se debe a que ambos son métodos de aprendizaje que utilizan los desarrolladores de software para perfeccionar los modelos de automatización de un determinado algoritmo, esto se traduce en que a través de estos métodos tienen como fin último la autonomía del sistema, tal como lo define Simon (2019). Así mismo, el autor resalta que la diferencia entre uno y otro método de desarrollo se da en la capacidad de aprendizaje y adaptación que la máquina es capaz de desarrollar y así mismo las especificaciones técnicas que se requieren para correr uno u otro programa. Vale la pena aclarar, que estas diferencias técnicas no generan conflicto para el presente trabajo por lo que no se discriminará al momento de referirse a ambas como "métodos de aprendizaje para el desarrollo de IIAA".

No obstante, aún no es clara la relevancia social de estas nuevas tecnologías más allá del miedo generalizado por el desplazamiento de la humanidad representado a través del arte. Con esto en mente, es relevante revisar el impacto y las proyecciones de crecimiento que tiene el mercado de las IIAA. De acuerdo con Peter Stone et al (2016), para el 2030 las IIAA estarán presentantes en mercados como: seguridad pública, educación, salud, infraestructura y educación, entre otros. De acuerdo con los datos publicados por Simon (2019), el

crecimiento del mercado de IIAA estima un crecimiento exponencial, pasando de generar 643.7 millones de dólares en ingresos a producir 36.8 billones de dolores para el 2025. Así mismo, se pronostica un potencial de crecimiento económico cerca al doble del percibido en 2016 para el 2035, para los países muestreados. En esta misma medida, vale la pena resaltar, que dentro del referido crecimiento industrial empresas como Google, IBM, Intel y Apple, entre otras, están compitiendo constantemente para adquirir startups o empresas dedicadas al desarrollo de IIAA. De la misma forma, el *McKinsey Global Institute* (2017) estimaba que estas compañías estarían invirtiendo entre 5 y 7 billones de dólares anualmente en el desarrollo de métodos de aprendizaje para IIAA, lo que representa el 56 por ciento de la inversión en tecnologías relacionadas con este campo.

A través de los anteriores indicadores se evidencia la relevancia que industrialmente ha adquirido este sector. No obstante, es necesario aterrizar la implementación de la inteligencia artificial al día a día social. Si bien actualmente se cuenta con un gran número de ejemplos, para Simon (2019) algunos de los más accesibles son los algoritmos de recomendación o sugerencia utilizados por plataformas de *streaming* con el fin de indicarle a sus usuarios que contenido deberían consumir en un futuro, basados en las series, películas o documentales que ya han visto. Así mismo, el autor describe al *Google Smart Reply* como servicio que permite automatizar respuestas en tiempo real a correos electrónicos tomando como ejemplo respuestas anteriores y adaptándolas al nuevo mensaje. De esta forma se demuestra, la exposición social a un gran número de estas tecnologías sin siquiera ser conscientes de esto, sin embargo, las aplicaciones de las IIAA no se han reducido a compañías especializadas en servicios de comunicación y/o entretenimiento.

Uno de los campos con mayor sensibilidad en los que las IIAA están siendo implementadas es la medicina. Por una parte, Challen et al (2019) demuestran como, en la actualidad, el mayor foco investigativo en las ciencias de la salud consiste en la aplicación de ML para generar diagnósticos, a partir de grandes cantidades de datos. Así mismo, sistemas de predicción están siendo utilizados, en EEUU, para determinar características sociales de los individuos que han ingresado al sistema de justicia penal de este país, y determinar la probabilidad de reincidencia de estas personas, lo anterior con el fin de determinar la posibilidad de acceder a beneficios judiciales, como el sistema COMPAS analizado por Propublica (2016).

Habiendo identificado apenas dos de las áreas sensibles en las que las IIAA están siendo implementadas, así como las proyecciones de crecimiento de este marcado, es claro que la identificación y entendimiento, y la consecuencial mitigación, de las externalidades negativas que podrían gestarse por la implantación de estas nuevas tecnologías es necesaria, sobre todo cuando se identifica el papel protagónico, y por lo tanto la expansividad, que las IIAA tendrán en el que hacer social cotidiano en el futuro. Evidentemente, estas características materializan muchísimas de las preocupaciones que sociólogos como Beck y Giddens expresaron en sus trabajos respecto a la nueva modernidad y a los riesgos que esta acarrea.

No obstante, es evidente que no cualquier avance tecnológico conlleva una serie de riesgos de tal relevancia que deba ser objeto de la intervención directa del Estado. Por lo tanto, con el fin de justificar o no una presencia intervencionista Estatal en la actividad de la referencia es necesario discriminar: a) el objeto sobre el cual opera la inteligencia artificial; y (b) el punto de la cadena de consumo desde el cual se genera el riesgo. Sobre el primer elemento, es claro que la categorización piramidal de riesgos que sigue el EC (2021b, 2021c) en su

propuesta es muy acertada, toda vez que identifica actividades sociales cuya automatización supondría una serie de riesgos de mayor relevancia, como la selección de candidatos aptos para un puesto de trabajo, y otras circunstancias frente a las que no valdría la pena intervenir, como los algoritmos de sugerencias en las plataformas de *streaming*. Con base en lo analizado, es evidente que de aprobar la intervención directa del Estado sobre el uso y/o desarrollo de IIAA, esta debería limitarse a las actividades cuyos riesgos se relacionen de una manera específica y clara con la seguridad y protección de las personas, así como con sus derechos fundamentales.

En consecuencia, a continuación, se expondrán casos en los que el desarrollo y/o uso de las IIAA generó resultados dañinos a las personas objeto del mismo y se analizarán sus causas.

3. Sobre las inteligencias artificiales y su impacto social

3.1. De lo general a lo particular: inteligencias artificiales y la toma de decisiones en casos concretos

3.1.1. Sesgo algorítmico y la catálisis de la desigualdad

Como se expuso anteriormente, las IIAA poseen una característica diferenciadora de otros sistemas de automatización o procesamiento de datos, la cual radica en su capacidad para aprender y adaptarse. Dicha facultad se logra a través de diversos procesos de aprendizaje aplicables a un algoritmo. Es así, como la Ciencia de Datos juega un papel determinante en la capacidad final del algoritmo de procesar datos nuevos o que no conocía. Usualmente, las IIAA pasan por un proceso de entrenamiento en el que analizan grandes cantidades de datos,

como lo explican Challen et al (2019). Vale la pena resaltar, que existen dos macro categorías en estos procesos de aprendizaje, el aprendizaje supervisado y el no supervisado. Como sus nombres lo indican, la diferencia sustancial entre uno y otro proceso de aprendizaje radica en el seguimiento que el desarrollador hace frente a los resultados de las pruebas ejecutadas. Luego de superar dicha etapa, el sistema es capaz de procesar datos nuevos y arrojar resultados confiables frente a los mismos, incluso aprendiendo de dichos datos, llegando a corregir o reforzar "conclusiones" a las que haya llegado durante su entrenamiento, tal como lo indican Challen et al (*ibidem*).

Con base en esto, es claro que la implementación de IIAA contribuirá en gran medida al avance en las diferentes áreas en que sean implementadas, sobre todo cuando se tiene en cuenta el impacto que las IIAA representan en la productividad. Sin embargo, existen límites claros frente a las capacidades de implementación de estas nuevas tecnologías. El sesgo algorítmico es, a nivel social, uno de los mayores retos que la automatización con base en ML y DL trae consigo misma. Dicho fenómeno es la materialización de error(es) estadístico(s) en el tratamiento de los datos de entrenamiento; de acuerdo con Leavy et al (2020) la consecuencia de este fenómeno se materializa en predicciones viciadas o irreales. Para Nelson (2019) Este riesgo se potencia en gran medida en sistemas que operan con datos sensibles, esto en la medida en que el procesamiento de este tipo de información debe tener en cuenta variables muy volátiles dentro del contexto en el que se dan, sin las cuales los datos puedes ser muy o nada valiosos.

Esto produce el ensanchamiento de las brechas de desigualdad social existentes y su justificación estadística. Los efectos de este fenómeno, percibidos por Silberg & Manyika (2019) van desde la desatención a una comunidad especifica basado en la estadística de su

consumo de medicamentos y como se relaciona esto a la gravedad de sus enfermedades, hasta la implementación de políticas públicas policivas agresivas, adoptabas con base en el reporte de delitos en un área específica de una ciudad. Claramente, el manejo de datos sensibles en el desarrollo de herramientas que apoyarán la toma de decisiones para el manejo de los fenómenos que generaron dichos datos no es un tema menor y por lo tanto debe ser gestionado con el objetivo de reducir en la mayor medida posible los efectos nocivos que un mal análisis estadístico puede generar.

3.1.2. Caso de estudio: COMPAS

En el campo de la justicia, sobre todo en el ámbito penal, dicho fenómeno ha sido foco de debate a partir de los análisis desarrollados frente a las decisiones tomadas con base en el COMPAS, el cual es uno de los softwares de evaluación de riesgo más utilizados por la justicia norteamericana, de acuerdo con ProPublica (2016a). Para el caso concreto la ONG realizó un estudio frente a las sentencias, y sus bases, impuestas en Broward County, Florida. De acuerdo con la investigación realizada, se logró evidenciar como una sobrerrepresentación de comunidades históricamente criminalizadas en los datos de entrenamiento del algoritmo generaba resultados poco confiables a la hora de determinar el riesgo que dichas personas representaban y, así mismo, como miembros de otras poblaciones que habían obtenido puntajes significativamente más bajos, con el tiempo reincidían en conductas similares u objetivamente más graves. Inclusive, se identificó como había subpoblaciones que se veían aún más desprotegidas ante el algoritmo de predicción, puntualmente las mujeres Afro. Esto se debía a que, sumado al problema de sobre representación anteriormente expuesto, el sistema COMPAS empleado no estaba diseñado

para analizar mujeres. Es decir, la *Data* de entrenamiento no contemplaba esta población. Vale la pena resaltar que la empresa creadora del COMPAS sí ofrecía un servicio especializado para mujeres, pero Broward County no lo empleaba.

Este análisis se adelantó tomando los datos de 11,757 personas que fueron juzgadas en una etapa previa al inicio del juicio oral. El software proyectaba, por lo menos, tres valores distintos cada uno relacionado con el "riesgo de reincidencia", "riesgo de violencia" y "riesgo de no presentación". Cada categoría era determinada con un puntaje del 1 al 10 donde en el rango de 1 a 4 se catalogaba como "bajo", 5 a 7 era "medio" y 8 a 10 era considerado "alto". Es importante señalar que de acuerdo con el estudio adelantado por los analistas de ProPublica (2016b) la definición de reincidencia fue fundamental para adelantar su análisis, por lo que debieron reconstruir el concepto procesado por el software, con el fin de obtener resultados mucho más objetivos. De esta manera agregaron y o excluyeron factores como: fecha de ocurrencia de los hechos "nuevos" con relación a la fecha del análisis en Compass y naturaleza de la nueva ofensa. Así mismo se apoyaron en estudios adelantados por parte del FBI y por la Comisión de sentencias de los EE. UU..

Fruto del análisis anteriormente citado se logró concluir que:

1. El software tuvo un porcentaje de falsos positivos, frente a la reincidencia, del 45% frente a personas negras, mientras que el porcentaje en sus homologas blancas fue del 23%.

2. El porcentaje de error con relación a los falsos negativos, en lo que coincide a la reincidencia, fue del 48% en personas blancas, mientras que en la población afro fue apenas del 28%.

3. Incluso al hacer análisis con otras variables como: crimenes anteriores, reincidencia en el futuro, edad y género, la probabilidad de asignación de un mayor nivel de riesgo en cabeza de una persona negra, en comparación al de una persona blanca, era del 45%.

4. Las personas negras tenían el doble de posibilidades de ser clasificadas erróneamente como reincidentes violentes. Así mismo, reincidentes violentos blancos tenían una probabilidad del 63% de haber sido clasificados erróneamente como sujetos de bajo riesgo, en comparación a sus homólogos negros.

5. El análisis con variables controladas frente a la reincidencia violenta indicó que las personas negras tenían una probabilidad del 77% de contar con una clasificación de riesgo mayor, que la de las personas blancas. (Propublica, 2016b. pág. 2)

Los sectores que abogan por el uso de puntajes de riesgo, y en consecuencia por las tecnologías que facilitan su cálculo, sostienen que su empleo contribuiría a la disminución de las tasas de encarcelamiento. De acuerdo con la información recolectada por ProPublica (2016a), a partir del 2005, tan solo tres años después del inicio de su empleo en Virginia, el crecimiento de la población carcelaria del estado decreció al 5% en comparación al 31% registrado en años anteriores. En esa medida, servidores como Mark Boessenecker, Juez Superior del condado de Napa para el 2016, han descrito estas tecnológicas como útiles. En jurisdicciones como las de Napa el Departamento de Subrogados utiliza puntajes de riesgo para proponer subrogados de la pena en casos individuales. Sin embargo, el juez indicó ante la ONG que dichos puntajes debían ser utilizados con precaución, toda vez que existirán variables no contempladas en la estructuración del software que generarían volatilidad en los resultados, como por ejemplo la situación laboral de la persona analizada.

A lo largo del territorio estadounidense la implementación de estos softwares ha ido creciendo de manera acelerada. De acuerdo con Propublica (2016a), para 2010 la mayoría del estado de New York, excepto New York City, hacia uso de esta herramienta, pero no fue hasta 2012 que publicaron los estudios de fiabilidad del programa. Ahora bien, de conformidad con el análisis realizado por ProPublica, el estudio adelantado por las autoridades no tuvo en cuenta el factor racial a la hora de presentar sus resultados, situación frente a la cual manifestaron que no era necesario debido a la falta de correlación entre el factor racial y la población sobre la cual se corría el programa.

Loomis vs Wisconsin: el sistema COMPAS, de la teoría a la práctica. Ahora bien, es evidente que no existe normatividad alguna que generé una obligación de obediencia de los jueces al resultado del programa. Sin embargo, existen casos en los que jueces han citado directamente los resultados del algoritmo para sustentar sus decisiones. El caso Loomis vs Wisconsin ha sido uno de los que más ha llamado la atención de los medios. De acuerdo con la Corte Suprema de Wisconsin (2016), en el citado caso Eric Loomis fue imputado con cinco cargos criminales por un supuesto tiroteo en La Crosse, frente a esto el defendido decidió allanarse a dos de los cinco cargos imputados. Finalmente, en la audiencia de lectura de fallo e individualización de la pena el Juzgado se basó en un análisis de riesgo realizado en COMPAS y determino la pena a cumplir en seis años de prisión y cinco de libertad condicional. Dicha decisión fue apelada por la defensa de Loomis bajó dos cargos principales, ambos argumentos se relacionaban con violaciones al debido proceso del sentenciado: el primer punto se centraba en la imposibilidad de defenderse frente al uso del algoritmo, toda vez que su estructura y por lo tanto los métodos algorítmicos empleados para llegar a los puntajes de riesgo estaban cobijados por leyes de propiedad industrial y por ende

no eran de acceso público y el segundo punto de apelación se basaba en que COMPAS tomaba en cuenta factores como la raza y el sexo como variables significativas para la predicción de riesgo, como se observa en el *Brief of Defendant-Appellant* del caso de la referencia.

Una vez analizado el recurso de apelación, la Corte Suprema de Wisconsin determinó que la solicitud del defendido no era procedente. Sin embargo, vale la pena hacer hincapié en la posición adoptada por los Jueces Bradley y Abrahamson,[5] quienes advirtieron que el uso de estas tecnológicas debía estar acompañado de un análisis mucho más profundo acerca de las razones de la decisión y así mismo los jueces debían ser informados de las limitaciones que el software poseía. Ahora bien, tal como se señala en el *Harvard Law Review* (2017), el pronunciamiento del órgano de cierre estatal hace poco en prevenir a los jueces frente a las fallas que presentan este tipo de herramientas y falla en considerar la capacidad de los jueces de comprender y analizar las características técnicas del programa. Sumado a esto, no hay un llamado de atención contundente a empresas como Northpoint a comparecer para aclarar las dudas alrededor de su producto.

Ahora bien, en este punto es evidente que la forma en la que se está implementando este tipo de tecnologías, sumado a los fallos técnicos identificados por ProPublica y a la falta de voluntad de los implementadores del software para determinar la falibilidad, han generado una serie de resultados dañinos en grupos sociales que históricamente han sido marginalizados, llegando a debilitar derechos fundamentales como la igualdad material, acceso material a la justicia, entre otros.

[5] Loomis, 881 N.W.2d at 769-70 and 774-75.

3.1.3. Proxy Data y sesgo: Caso Apple Card

A finales de 2019, una serie de denuncias sobre conductas discriminatorias por parte del sistema *Apple Card* derivaron en una investigación por parte del Departamento de Servicios Financieros de Nueva York sobre el mencionado producto (NYDFS). De acuerdo con lo investigado por O´Sullivan, L. (2021), el proceso inicio luego de la viralización de un tweet, que exigía claridad respecto a los motivos generadores de los topes máximos relativos a los créditos financieros que Apple y Goldman´s Sachs ofertaban. La disputa inició luego de que el usuario identificará que la diferencia de montos entre los topes máximos ofertados a él y los ofertados a su pareja discernía de manera desproporcionada, teniendo en cuenta que tanto el cómo su esposa compartía la misma capacidad financiera, tal como lo explica Vigdor (2019).

De acuerdo con la investigación adelanta por el NYDFS (2021) se logró determinar que los métodos utilizados por los investigados, en este caso un algoritmo de procesamiento de *Data* no violaba la normativa desarrollada por el *Equal Credit Opportunity Act* (ECOA) de 1974. El NYDFS llegó a esta conclusión luego de revisar, entre otros elementos, que el algoritmo no utilizaba "características discriminatorias", tales como: etnia, genero, estado civil, entre otras. Sumado a lo planteado, también se desarrollaron *flip tests* o pruebas de cambio, en las cuales se alteraban datos en las aplicaciones a créditos y se analizaba la variación entre el monto ofertado con los datos originales y el monto ofertado tras el cambio de variables. Puntalmente se tomaron datos de 400,000 aplicantes en Nueva York y se aplico el metodo de análisis anteriormente mencionado, con control sobre las nombradas "características discriminatorias". De acuerdo con lo informado, el resultado no presentó

variaciones significativas entre las ofertas de crédito financiero, por lo que se llego a la conclusión anteriormente mencionada.

Sin embargo, análisis posteriores evidenciaron que la NYDFS había dejado de considerar el uso de *proxy Data* dentro de los modelos empleados para la oferta de créditos financieros. De acuerdo con O´Sullivan (2021), el uso de datos estadísticamente relacionados con grupos demográficos específicos, o *proxy Data*, se ha convertido en una manera de perfilar a las personas por su estatus social, sexo, etnia, afiliación política, etc., sin asumir el uso directo de estos datos en los modelos algorítmicos empelados. Este fenómeno permite el desarrollo de modelos altamente discriminatorios que pueden ser empleados de manera libre con poco o nada de repercusiones para sus usuarios; lo expuesto se debe, tal como lo plantean Barocas & Selbst (2016), a la dispareja velocidad en la que la Ciencia de Datos ha evolucionado en comparación a las fuentes normativas que regulan las diferentes actividades sociales en la que esta disciplina está siendo empleada. Exactamente, esta es la situación evidenciada en el caso de la referencia, toda vez que debido a la antigüedad del ECOA esta no consideraba el impacto y uso de *proxy Data* como fuente de discriminación, es en esa medida que el sesgo algorítmico fue pasado por alto.

3.1.4. Consideraciones

En este punto se evidencia la gran relevancia de algunos de los sectores en los que las IIAA están siendo empleadas sin control de calidad alguno. No solo se encuentran casos relacionados con el sector financiero o el aparato de justicia, sino también con la prestación de servicios de salud, como lo analizan Obermeyer et al. (2019), donde se ha demostrado que las IIAA empleadas para la predicción de riesgos de salud, y por lo tanto la accesibilidad a

seguros y servicios, han sido desarrolladas con datos que generan sesgos raciales. Con base en lo analizado, se evidencia que la existencia de determinados sesgos en los algoritmos de procesamiento de datos empleados en escenarios sensibles puede afectar la prestación efectiva de servicios esenciales en la sociedad, y por lo tanto podría considerarse de alto riesgo. Es en esa medida, que la presencia de factores que alimenten sesgos nocivos en los algoritmos empleados para la automatización de tareas en entornos sensibles, como los analizados, representa un riesgo sobre el cual debe haber parámetros claros de manejo y administración. Así mismo, y a través de los casos de la referencia se puede apreciar como la falta de normativa actualizada e informada genera vacíos legales que pueden ser aprovechados para la evasión de responsabilidad en estos casos.

Sumado a esto, también es relevante resaltar la íntima relación que tienen los sistemas de automatización y la Ciencia de Datos, toda vez que, como se ha demostrado, el empleo desregulado de ambas ciencias representa un espacio idóneo para la agudización de riesgos de fallo en la prestación de servicios. Lo explicado cobra mayor relevancia cuando se considera que dichos riesgos pueden verse materializados en actividades en las que se opera con derechos fundamentales de las personas. Por lo tanto, la reglamentación que se ocupe de estas actividades deberá tomar en cuenta los diferentes campos científicos relacionados con desarrollo de herramientas de automatización y deberá ser capaz de reconocer los diferentes mecanismos de administración de riesgos aplicables a cada uno de manera particular.

Así mismo, tal como se evidencia en el caso de Broward County, estos estándares de calidad no le serán exigibles únicamente a la compañía desarrolladora del sistema, sino también a cualquier persona que este encargo de operarlo. Por lo tanto, es claro que a la hora de desarrollar un marco jurídico reglamentario frente al uso y/o desarrollo de las IIAA, será

necesario especificar sobre quien recae la obligación de cuidado; en esa medida se deberá tener como criterio orientador el reconocimiento de la naturaleza de los diferentes riesgos generados a lo largo de la cadena de consumo de estas tecnologías, facilitando así una eventual imputación de responsabilidad.

Otro punto que vale la pena resaltar a partir de los casos de la referencia, radica en la existencia de protección frente a la propiedad intelectual de una empresa; elemento que podría llegar a entorpecer el control social y estatal sobre los métodos empleados para el desarrollo de las IIAA. Es evidente que esta situación será extensible a situaciones homologas, por lo tanto, la formulación de un marco jurídico que intervenga directamente estos asuntos debe considerar la relevancia de esta variable.

Finalmente, es importante tener en cuenta que, a grandes rasgos, para el presenté análisis es irrelevante la evaluación del elemento subjetivo de las conductas desplegadas por las personas involucradas en los casos de la referencia, por lo menos no en esta etapa de la investigación. Lo señalado parte de la demostración de relevancia de este tipo de conductas en los escenarios analizados, lo que implica que incluso el actuar culposo podrá ser fuente de responsabilidad.

3.2.De lo particular a lo general: la implementación de inteligencias artificiales en contextos de impacto generalizado

3.2.1. Retos democráticos: micro-targeting e Inteligencias Artificiales

Si bien los riesgos analizados en los casos anteriormente expuestos podrían ser catalogados como individuales, no debe olvidarse que las IIAA están siendo implementadas

en otros campos en los que su uso ha conducido a la optimización de procesos complejos como la construcción de campañas políticas más eficientes, llegando a tener impactos a escala local, regional o internacional. Dentro de este escenario un caso de gran relevancia es el *Cambridge Analytica scandal* (CAS). Es de conocimiento público que dicho suceso recibió una cantidad relevante de atención mediática y gubernamental relativa a la protección de datos de los usuarios y a la transparencia de los procesos democráticos; también es cierto que el elemento tecnológico fue poco analizado.

De acuerdo con Dubber et al. (2020), uno de los elementos centrales que le permitieron a *Cambridge Analytica* (CA) tener el impacto que tuvo en Estados Unidos (EEUU) y en el Reino Unido (RU) fue el uso de IIAA para el procesamiento de datos y la distribución de contenido en plataformas digitales, a través de la lógica de *micro-targeting*. Tal como lo explican Tapper et al. (2006), el *micro-targeting* consiste en realizar un estudio de los patrones conductuales de un grupo especifico de personas con el fin de determinar elementos o *proxys* representativos de su comunidad. Sumado a lo planteado, los autores indican que a través de esta herramienta es posible generar un discurso mucho más directo y personal hacia el público objetivo, lo que en política implica una mayor tasa de identificabilidad entre el elector y el discurso político al que lo exponen.

Si bien el CAS data del 2018, e investigaciones como la adelantada por Confessore (2018) han demostrado que los datos utilizados para generar los modelos habían sido extraídos desde el 2014, analistas políticos han empleado estas practicas desde mucho antes de la participación de CA. Por ejemplo, a partir de lo trabajado por Tapper, et al (2006) se logró determinar que en 2004 la campaña presidencial de Bush identifico personas influyentes, conocidos ahora como *influencers*, y trabajó fuertemente por transmitir su discurso político

a través de ellos, toda vez que ese medio de difusión era mucho más efectivo. Si bien la práctica de *micro-targeting* es altamente cuestionable desde un punto de vista ético, es aún más preocupante que este tipo de herramientas de manipulación puedan ser explotadas de manera libre a través de procesos de automatización.

Lo expuesto cobra mayor relevancia aún tras un análisis de las respuestas gubernamentales adoptadas tras CAS. Tanto USA como RU ordenaron el pago de altas sumas de dinero a titulo de multa por las violaciones a la privacidad de los datos de los usuarios efectuadas por Facebook y CA, como lo expuso Mcauley (2020). Es claro, que a partir de este incidente se adoptó una perspectiva mucho más proteccionista frente a la recolección y manejo de *Big Data* por parte de diferentes legislaciones. Sin embargo, debido a la nula respuesta frente al uso indebido de las IIAA en el caso de la referencia, no se logró sentar ningún precedente frente a su uso dentro de contextos de interés general, como lo sería la política.

3.2.2. Deepfake

A partir de 2017 el termino *Deepfake* (DF) registró su primer uso "viral". Chadha, et al (2020) explica como esto sucedió luego de que un usuario de Reddit publicara videos íntimos en los que supuestamente se apreciaba la participación de una celebridad. Luego de una investigación se determinó el uso de IIAA para la generación del video alterado. La tecnología empleada para esta tarea empezó a ser descrita como *Deepfake*, una mezcla de los términos *Deep Learning*, analizado anteriormente, y *fake*, o falso, por lo que nos encontraríamos frente a una falsificación a través del DL. Este tipo de IIAA analizan diferentes componentes comunicativos de una persona como su voz, expresiones faciales, manierismos, etc., y es capaz de recrearlos a través de *image mapping* y otras tecnologías.

Evidentemente, el uso de este tipo de tecnologías tiene tanto elementos positivos como negativos. Por ejemplo, y tal como lo plantea Wester (2019) desde la industria del entretenimiento, como el cine, las aplicaciones del DF van desde la actualización de material antiguo, hasta la posibilidad de finalizar la producción de películas y series sin necesitar la presencia de los actores. Así mismo, el autor plantea como las industrias electrónicas y de marketing podrían darle uso a este tipo de tecnologías con el fin de generar productos mucho más amigables en entornos digitales, generando así una mejor experiencia de usuario. Por otra parte, y de acuerdo con Chadha, et al (2020), el principal riesgo del uso de *DF* es la capacidad de este de transmitir desinformación o noticias falsas con la mayor eficiencia posible. Este temor es compartido tanto por el sector público como privado, toda vez que, con el ritmo de desarrollo técnico registrado, su uso en tiempo real es una preocupación a gran escala, tal como lo indican ambos autores.

Werner (2020) realiza un trabajo extensivo en la identificación de casos reales en el que se ha registrado el uso de *DF*, es importante resaltar que el autor logró identificar diferentes tipos de fraudes personales en los que se ha conseguido suplantar la identidad de ejecutivos de alto nivel, a través de llamadas telefónicas, para generar transferencias bancarias a cuentas de terceros. Otros casos para tener en cuenta, con resultados a escala representativa son los expuestos por Harwell (2019): a) la reparación de Ali Bongo del 2018; y b) la confesión de un ciudadano en Malasia de haber tenido relaciones sexuales con un miembro del gabinete de ministros. El impacto del primer caso se evidenció a través del intento de golpe de estado militar desarrollado en Gabón posterior a la publicación del video, en el segundo caso una serie de protestas en contra del gobierno se desarrollaron en el territorio malasio luego de que la opinión pública considerara como ciertas las alegaciones.

Es así, que se ha logrado identificar el crecimiento de *DF* de "fuente abierta", esto implica que el código fuente de la IIAA necesaria para desarrollar el *DF* se encuentra de manera gratuita y pública en el internet. Vale la pena resaltar que existen diferentes tipos de *DF*, como lo exponen Chadha, et al (2020) y Wester (2019), con diferentes usos y capaces de emular diferentes elementos de una persona. Como respuesta a este fenómeno las ciencias forenses han desarrollado métodos de análisis digitales que permiten la identificación de las *DF*, tal como lo exponen los autores en sus respectivas investigaciones. Por lo tanto, el control sobre la información se convierte en una cuestión de acción y reacción, evidentemente si el DF es empleado de manera organizada contrarrestar sus efectos puede llegar a ser mucho más complejo que únicamente comprobar la manipulación de la imagen y/o video, como se evidencia en los casos de estafa.

A partir de esto se demuestra la capacidad de movilización generada a través del empleo de DF, por lo que no es sorpresivo que algunas naciones consideren estas IIAA un tema se seguridad nacional. Algunos de los ejemplos que vale la pena resaltar son *la U.S. National Defense Authorization Act* (NDAA) del 2021 y la *Identifying Outputs of Generative Adversarial Networks Act* (IOGANA) del 2019, ambas fuentes normativas tienen como fin impulsar la investigación, así como el análisis de riesgo, relacionado con el DF en EEUU. Adicionalmente, Pradhan (2020) ha identificado una tendencia encaminada hacia la prohibición de estas tecnologías en contextos políticos en algunos estados. Por otra parte, otros cuerpos de gobierno como la UE y China han trabajado en la implantación de cuerpos normativos que apuntan a la regulación de este tipo de tecnologías. Puntualmente se está a la espera de los resultados de la legislación china, toda vez que esta fue implementada apenas el pasado 11 de enero de 2023. Así mismo, es importante resaltar que la propuesta de

regulación de IIAA de EU contempla la posibilidad de permitir el uso de tecnologías *DF*, con el cumplimiento de una serie de requisitos mínimos, los cuales se enfocan en la transparencia de los desarrolladores, como lo explican Van Huijsteee, et al. (2021).

Lo anterior cobra aún mayor relevancia al analizar dos de los cuatro grupos desarrolladores de DF identificados por Wester (2019). En primer lugar, se tiene a grupos estatales dedicados a desarrollar este tipo de tecnologías con el fin de intervenir en asuntos oficiales de otras Naciones, o en la propia, y grupos de delincuencia organizada. Evidentemente, en estos casos la prohibición de uso de este tipo de tecnologías no generaría una disminución de empleo, toda vez que las actividades para las que son empelados se dan y seguirán dando desde la ilegalidad. Sin embargo, la creación de obligaciones especiales para los desarrolladores legales de esta tecnología, frente a la comercialización de está, permitiría una mejor respuesta por parte del Estado frente al uso ilegal de la misma, toda vez que podría identificar con mayor seguridad las características técnicas empleadas en los programas, lo cual permitiría mayor velocidad de reacción. Así mismo, la consideración del DF como una herramienta cuyo uso debería estar regulado permitiría la intervención estatal a fin de controlar los riesgos de su empleo, así mismo se podrían generar modelos de colaboración entre los sectores informativos, tanto públicos como privados, para implementar mecanismos de filtro que prevengan la introducción de productos DF en sus plataformas.

3.2.3. Consideraciones

Ahora bien, evidentemente en los casos de *micro-targeting* y *DF*, a diferencia de los casos analizados bajo los numerales 2.1 y siguientes, los riesgos concretados, y por lo tanto los daños generados, no surgen a partir de un manejo sesgado de los datos. Al contrario, el

funcionamiento de las IIAA empleadas fue el esperado, y tuvo éxito en su implementación. En esa mediada, la pregunta que se deriva de los hechos analizados es ¿de qué manera debería intervenir el Estado frente al uso consciente de tecnologías potencialmente dañinas? Esta pregunta inicia un nuevo debate, en la medida en que una respuesta positiva podría llevar a regular el comercio de este tipo de herramientas, con el fin de evitar su utilización en contextos sociales sensibles, como lo pueden ser la toma de decisiones políticas, lo que generaría una nueva carga de responsabilidad en el vendedor del algoritmo. Por otra parte, una respuesta negativa derivaría en que el vendedor estaría exento de responsabilidades de control sobre sus compradores, y por lo tanto únicamente el usuario podría ser objeto de reproche tras un mal uso de la tecnología.

Para responder el debate en cuestión sería pertinente analizar la metodología empleada para resolver esta temática en otras áreas. Un ejemplo útil podría ser la determinación Estatal de regular el mercado de armas de fuego, pero no el de objetos cortopunzantes o contundentes. Evidentemente, esta cuestión podría ser abordada desde diferentes perspectivas.

De acuerdo con Blocher & Reva (2021), la regulación al porte y/o comercio de armas se basa en la intensión del Estado de "(…) preservar la paz y proteger a los ciudadanos de amenazas armadas (…)[6]" (163). Siguiendo una línea similar, Siegel & Blocher (2020) concluyen que los principales intereses que fundamentan la regulación a las armas de fuego se encuadran en el marco de libre ejercicio de las libertades (movilidad, culto, expresión, entre otras); para estos autores las armas representan una fuente de peligro, que impide la

[6] Traducción propia.

libertad de goce de las libertades anteriormente señaladas, por parte de los sujetos no armados. Sumado a esto, es importante señalar como en algunos de los Estados que cuentan con regulaciones sobre armas la discusión toma un tinte constitucional; esto se puede evidenciar en casos como: *Columbia v. Heller*, citado por Blocher & Reva (2021), en EE. UU., y la C-296 de 1995, en Colombia, sobre esta ultima se resalta:

> (…) según las estadísticas existentes, es posible sostener que el porte de armas promueve la violencia, agrava las consecuencias de los enfrentamientos sociales e introduce un factor de desigualdad en las relaciones entre particulares que no pocas veces es utilizado para fortalecer poderes económicos, políticos o sociales, Por eso los permisos para el porte de armas sólo pueden tener lugar en casos excepcionales. Esto es, cuando se hayan descartado todas las demás posibilidades de defensa legítima que el ordenamiento jurídico contempla para los ciudadanos. (Corte Constitucional, 1995).

Es claro que habrá quienes aboguen que la regulación responde al deseo estatal por mantener el monopolio de la fuerza en un territorio determinado, así como también podría sostenerse que la regulación responde al deseo del Estado por disminuir los índices de violencia, entre otros. Sin embargo, cualquier postura adoptada apuntara a que la regulación responde al interés del gobierno por administrar una fuente potencial de riesgo. Sumado a esto, se puede evidenciar que el riesgo no es mínimo, sino que tiene un impacto significativo con intereses estatales fundamentales, como la paz y la seguridad. Es en esa medida, que el Estado considera legitima la intervención en este mercado preciso, a diferencia de otros objetos que podrían ser utilizados como armas (cuchillos, navajas, bates, crucetas, etc.).

Es interesante ver cómo, para Cook et al. (1995), la regulación sobre los *arm dealers* corresponde a una medida encaminada a dificultar el acceso a armas de fuego a grupos determinados de personas, toda vez que se responsabiliza al comerciante de realizar una serie de evaluaciones para comprobar que, por lo menos preliminarmente, la persona que vaya a adquirir un arma de fuego no represente un peligro para la sociedad. Esta es medida que claramente podría ser implementa en el mercado de las IIAA y podría llegar a tener un impacto similar. Sobre todo, teniendo en cuenta que, como se ha demostrado hasta ahora, el desarrollo e implementación de estas herramientas tecnológicas reporta un gran número de ventajas para sus usuarios y la prohibición completa de las mismas representaría una gran perdida para la sociedad.

Ahora bien, la regulación también podría incluir: (a) la creación de un sistema de control sobre los compradores y el tipo de tecnología que estos adquieren; y (b) la obligación, por parte de los desarrolladores de registrar nuevas tecnologías y la forma en que operan, ante las autoridades correspondientes. Frente a la última opción, nuevamente, existirán objeciones desde una visión de secreto industrial y derechos de autor; por lo mismo, la implementación de este tipo de sistemas debe operar bajo estricta reserva, teniendo en cuenta también que podría llegar a comprometer intereses nacionales.

3.3.¿Un costo necesario? Inteligencias artificiales y exclusión social

3.3.1. Brecha digital

El concepto de brecha digital viene siendo utilizado desde la segunda mitad del siglo XX, tal como lo expone Lombana (2018). Este concepto, originariamente, hacía referencia a la

desigualdad que parte de la población sufría frente al acceso material a nuevos tipos de tecnologías, primero en Francia y posteriormente en EE. UU. (Romero, 2020). Ahora bien, actualmente la evolución del concepto ha llevado al empleo de términos como "desigualdades digitales" o "niveles" de la brecha. (Lombana, 2018: 5). Atendiendo la anterior idea, es importante reconocer que implica cada uno de los "niveles":

(a) Primer nivel: desigualdad frente al acceso a la tecnología;
(b) Segundo nivel: desigualdad frente a habilidades y prácticas que generan mejor participación, creación y circulación de contenido en línea; y
(c) Tercer nivel: desigualdad frente al aprovechamiento de los beneficios derivados del uso de tecnología. (Lombana, 2018)

En relación con lo anterior, un concepto que vale la pena recalcar es el de "brecha cognitiva", desarrollado por Crovi (2010). Según la autora, el fenómeno de brecha digital ha sido analizado desde una perspectiva determinista y poco integral. En esa medida, considera que elementos como el cognitivo deben ser abordados con el fin de comprender la situación y plantear e implementar soluciones pertinentes. Al comparar esta categoría con los niveles propuestos por Lombana, puede observarse que la autora caracteriza la "brecha cognitiva" como una mixtura de los dos últimos niveles de brecha digital, toda vez que la descompone en "al menos dos aspectos fundamentales: las competencias informáticas y el capital cultural." (Crovi, 2010: 11). El primer elemento se relaciona con la capacidad que un sujeto tiene de relacionarse técnicamente con la tecnología; es decir, sus competencias informáticas determinarán cual será el "control" que logrará ejercer sobre el sistema. Por otra parte, del segundo elemento dependerá la posibilidad que tendrá el usuario de "gestionar información y crear conocimiento" (*ibidem*).

Con base en lo analizado, debe resaltarse que dichos niveles se interrelacionan entre sí, pero no dependen enteramente unos de otros. Por ejemplo, superar la brecha de primer nivel (sucede en la mayoría de las naciones del Norte Global), no implica un aprovechamiento eficiente de las nuevas tecnologías, tal como lo explican Ocaña-Fernández et al. (2019). No obstante, no debe dejarse de lado que esto únicamente sucede en estos países específicamente, esto claramente se refleja a través de datos como los recogidos por el Banco Mundial, en los que se evidencia que, para el 2020, el 60% de la población mundial hacía uso de internet[7]. En esa medida, si bien el mero acceso a la tecnología no es un problema de las dimensiones que podía ser hace unas décadas, este no debe ser descartado, tal como lo plantea Lombana (2018).

Es así, que el concepto de Brecha Digital congrega un mayor número de subconceptos; los cuales, apuntan al potenciamiento de disparidades sociales relacionadas con la capacidad de acceso y relación con la tecnología.

3.3.2. De la incapacidad de aprovechamiento a la exclusión social

Ahora, teniendo en cuenta el fenómeno anteriormente reseñado se deben analizar sus impactos. Autores como Arévalo-Ascanio et al. (2015) y Lombana (2018) realizan observaciones relevantes frente al nivel de competitividad que las tecnologías generan a nivel micro y macroeconómico, respectivamente. Resulta de especial relevancia para el análisis de esta tesis tomar los casos reseñados por Lombana, de acuerdo con la investigación adelantada por el autor la diferencia de crecimiento económico entre China y Latino América, para 2035, será de 21.6% a favor de la potencia asiática. Así mismo, el autor advierte sobre la

[7] Vid: https://data.worldbank.org/indicator/IT.NET.USER.ZS

dependencia tecnología de otras naciones sus impactos negativos, dentro de los cuales vale la pena resaltar: (a) la seguridad de los datos de relevancia nacional; y (b) la ineficacia de los sistemas para procesar información de contextos propios de las naciones adquirientes. Por otra parte, Arévalo-Ascanio et al. (2015) advierten como la falta de cultura empresarial alrededor de la tecnología es una de las principales causas de subdesarrollo económico para pequeñas y medianas empresas; ahora, si bien la toma de muestras de este trabajo se centra en una población muy específica de sujetos, no es difícil identificar como estas conclusiones son extrapolables al resto del mundo.

Lo anterior, caracteriza un primer tipo de exclusión. Una exclusión económica-productiva, en la que aquellos miembros del mercado, sobre todo en el área de la producción, que no sean capaces de adquirir e incorporar las nuevas tecnologías a sus procesos, serán excluidos paulatinamente por competidores mucho más fuertes. Ahora, evidentemente el anterior argumento podría ser rebatido sosteniendo que esta es una dinámica básica de los mercados mundiales. No obstante, es importante considerar que el uso/aprovechamiento de las tecnologías no es enteramente potestativo, sino que viene condicionado por una serie de elementos que no dependen de los actores, sino más de las características del entorno en el que se han desarrollado. Sobre esto, Ocaña-Fernandez et al. (2019) sostienen que existe una necesidad por generar modelos de "alfabetización digital", enfocado en el desarrollo de habilidades blandas y duras encaminadas a que las personas logren aprehender de manera integral las funcionalidades de las nuevas tecnologías de la información y la comunicación. Esto, en la medida en que esta es lo única forma de iniciar un cierre de la "Brecha Cognitiva".

Ahora bien, reducir el impacto de la brecha digital a sus efectos sobre el mercado sería reduccionista; toda vez que, como lo indica Romero (2020), el acceso a la tecnología y sobre todo al internet:

> […] dan lugar a que la brecha digital se manifieste en diferentes planos o dimensiones: en el acceso a la sociedad de la información, si se considera la red como una macro biblioteca virtual en la que se encuentra todo tipo de información; en el acceso al comercio electrónico, el *e-commerce*; en el acceso a la formación, el *e-learning*; en el acceso a la administración electrónica, la *e-administration*. Este último uso de la red adquiere una relevancia superior ya que ofrece al ciudadano prestaciones y servicios, en ocasiones, imprescindibles para ejercitar derechos y deberes ciudadanos. (Romero, 2020: 5)

Es necesario hacer hincapié, sobre todo, en el último apartado del párrafo anterior. Tomando a Colombia, por ejemplo, se puede evidenciar una tendencia por la digitalización y el aprovechamiento de tecnologías para el ejercicio de diferentes derechos o el contacto con las entidades públicas. Ejemplos de lo enunciado son:

(a) los portales virtuales de radicación de denuncias o tutelas;

(b) la carpeta ciudadana digital;

(c) la digitalización de la Cedula de Ciudadanía;

(d) la implementación de Chatbots por parte de las entidades públicas para atender PQRS; y

(e) la posibilidad de gestionar pagos tributarios a través de oficinas virtuales de las respectivas entidades recaudadoras, entre otras.

En este punto, la brecha digital se convierte en un freno de bienestar social, toda vez que la imposibilidad de interacción eficiente con la tecnología limita la capacidad que las

personas tienen de ejercer y gozar de los derechos que les son reconocidos. Esta situación, se presenta con mayor intensidad en la población de personas de mayor edad, tal como lo presenta Romero (2020). No obstante, Ocaña-Fernandez et al. (2019) que algunos de los llamados "nativos digitales" tampoco cuentan con las habilidades necesarias para aprovechar al máximo lo que las nuevas tecnologías tienen para aportar. En esa medida, la participación ciudadana y libre ejercicio de diferentes derechos se ve limitado en la medida en que las personas cuenten o no con las capacidades para hacer parte de un mundo cada vez más digitalizado, tal como se ha venido explicando.

Inteligencias artificiales y brecha digital. En este caso el daño analizado no se genera por falla o uso ilícito de las IIAA, sino por la complejidad que su uso eficiente representa para los usuarios "analfabetos". Ocaña-Fernandez et al. (2019) indica, acertadamente, que la implementación de IIAA los diferentes ámbitos sociales generan una mayor demanda de conocimiento en los usuarios de los sistemas, toda vez que, para entender el funcionamiento de las IIAA, y así poder aprovecharlas, es necesario de un manejo mucho más técnico y preciso de las lógicas de operación de estos sistemas, que las necesarias para el uso de sistemas más simples.

3.3.3. Consideraciones

Frente al fenómeno anteriormente descrito, tal como se indicó, la doctrina propone nuevos modelos educativos con el fin de cerrar la brecha generada. Ahora bien, en el caso puntual de Ocaña-Fernández et al. (2019) los autores indican que las competencias digitales deberían ser desarrolladas a través de educación universitaria. Se considera relevante advertir que, si bien este sería un escenario ideal para alcanzar parte de la población, no sería una solución eficiente frente a grupos sociales que no acceden a educación superior. Por ejemplo, de

acuerdo con Statista (2022), en el 2020, únicamente el 41% de hombres y el 36% de mujeres, a nivel mundial, habían logrado acceder a estudios de educación superior (universitaria o técnica). Por lo tanto, resulta necesario considerar al grueso de la población que no cuenta con acceso a este nivel educativo, para generar políticas públicas eficientes de "alfabetización digital". No sería posible frenar el ensanchamiento de la brecha digital ya si existente, si excluyen grupos como las personas mayores, analizadas por Romero (2020). Es necesario hacer hincapié en que este no sería un problema reservado para países en vía de desarrollo; por ejemplo, en Alemania y de acuerdo con el *Statisches Bundesamt* (2020), únicamente el 18,5% de la población alemana logra obtener un título de nivel terciario.

Es así, que se evidencia que le existencia y ensanchamiento de la brecha digital representa un riesgo de alta relevancia, toda vez que cataliza diferentes tipo de desigualdades estructurales, ya analizadas, y termina por excluir a aquellos sujetos no alfabetizados digitalmente de la vida a través de la denegación del ejercicio pleno de diferentes derechos. Sobre todo, debe entenderse que, por el carácter progresivo de los derechos políticos, civiles, económicos, culturales y sociales, la implementación de medidas que restrinja su goce podría ser considerada como fuente de responsabilidad en un escenario internacional.

Con base en lo analizado, una política encaminada a cerrar, o por lo menos frenar, la brecha digital generada por la implementación de IIAA a través de programas universitarios, no sería suficiente y únicamente generaría una acumulación de capital cultural en aquellos sujetos capaces de acceder al nivel terciario educativo.

Adicionalmente, el análisis desarrollado frente al fenómeno de la "Brecha digital", permite categorizar dos escenarios de exclusión:

(a) el primero, individual, centrado en las personas (jurídicas o naturales; comerciantes o consumidores; ciudadanos o inmigrantes); y

(b) una exclusión estatal, esta última generada a partir de la incapacidad de participación en las nuevas dinámicas de la sociedad internacional.

Si bien ambas formas de exclusión tienen causas similares, se evidencia que los impactos de estas varían de una forma a otra. Por lo tanto, la lógica detrás del trabajo social necesario para mitigar efectos en unos y otros debería asumir estas diferencias. Esto cobra sentido, al comprender que las políticas públicas encaminadas a "alfabetizar digitalmente" a un agente del mercado que evidencia la forma en que la competencia está logrando excluirlo a través de la implementación de nuevas tecnologías no será la misma que se requerirá para generar cambios en la perspectiva que una persona, de la tercera edad, tiene frente a la necesidad de "alfabetizarse digitalmente".

En adición, resulta relevante resaltar las observaciones relativas a una especie de "soberanía informática", término comparable con el de soberanía alimentaria, que es introducido por Lombana (2018). De acuerdo con Nyéléni (2007) la soberanía alimentaria consiste en la capacidad/derecho de cada persona de autogestionar todo lo relativo a su alimentación, sin la incidencia de terceros partidos. De esta manera, nos damos cuenta de que Lombana (2018) tiene una idea muy similar en mente, cuando nos advierte de los peligros derivados de una dependencia tecnológica frente a economías extranjeras. Debe convertirse en una prioridad alcanzar grados de independencia/soberanía significativa en los campos informáticos, sobre todo cuando estos están empezando a ser operados a través de IIAA.

Ahora bien, como se ha demostrado anteriormente, la implementación de estas tecnologías si bien es paulatina no se verá frenada, por lo menos no por ahora. En esa medida, un riesgo

de gran envergadura ligado al desarrollo e implementación de inteligencias artificiales en las actividades sociales que desarrollamos consiste en la ampliación de la ya existente brecha digital. Por lo tanto, la caracterizada "alfabetización digital" se convierte en una necesidad generalizada para el grueso de la población. Sobre todo, cuando se pueden identificar tendencias mundiales hacía el desarrollo e implementación de nuevas tecnologías en casi todos los ámbitos de interacción social, como se ha demostrado en capítulos anteriores.

4. Medidas estatales respecto a las inteligencias artificiales

4.1.Inteligencias artificiales en la Sociedad Internacional

De acuerdo con OECD.AI (2021a) se tienen datos de más de 800 políticas públicas encaminadas al desarrollo y análisis de las IIAA en 69 países. Según los datos registrados, USA es el país con mayor numero de iniciativas (77), seguido de la Unión Europea (59) y el Reino Unido (57). Vale la pena aclarar que estos datos no reflejan, necesariamente, la voluntad por regular directamente las actividades económicas relacionadas con las IIAA. Otro hito relevante fue la adopción, por parte de los 193 miembros de la UNESCO, de una serie de principios respecto de "Ethics of Artificial Intelligence". De acuerdo con lo planteado por UN NEWS (2021) "el texto apunta a resaltar los beneficios de las inteligencias artificiales, mientras reduce los riesgos que estas conllevan[8]".

Actualmente existen dos casos particulares que vale la pena analizar: China y la Unión Europea. De acuerdo con OECD.AI (2021b) el *National New Generation Plan* de China

[8] Traducción propia.

apunta a convertir a China en el principal centro de innovación a nivel mundial respecto a desarrollo de IIAA, para 2030. Con este fin en mente, el gobierno chino apunta a construir una economía de 130 mil millones de euros alrededor de este tipo de tecnologías. Es importante resaltar que, actualmente, China cuenta con regulación respecto a algoritmos de recomendación de contenido, como lo explican Sheehan & Du (2022). De esta forma es que se apunta a controlar la forma en que la información es difundida en China. La normativa de la referencia busca generar obligaciones de registro por parte de los usuarios de la tecnología, con el fin de identificar la forma en que esta opera. Principalmente, se busca recolectar información respecto al funcionamiento de este tipo de algoritmos. De acuerdo con los autores, lo anterior atiende a una política a mediano y largo plazo a través de la cual, el gobierno chino, espera tener un entendimiento mucho más profundo de la forma en que las IIAA operan. Esto, permitiría obtener el *know-how* necesario para impulsar el avance hacia las metas planteadas para 2030.

4.2.“The Artificial Intelligence Act”

En abril de 2021 el parlamento y el Consejo europeo presentaron una propuesta de regulación sobre IIAA aplicable a la totalidad de los países miembros de la Unión. El citado documento consta de cuatro secciones diferentes, tal como lo indica la Comisión Europea (2021b). A grandes rasgos la Comisión (2021a) identifica seis tipos de riesgos/daños relacionados con las inteligencias artificiales, entre los cuales vale la pena resaltar el impacto relacionado con la seguridad de las personas y el ejercicio de sus derechos humanos (8). Así mismo, indica que el modelo regulatorio deberá basarse en un análisis de riesgos y, en esa

medida, únicamente serán las IIAA que representen altos riesgos[9] las que deberán ser altamente reguladas.

A partir de lo expuesto, es importante resaltar que la Comisión (2021a) reconoce que la nueva regulación deberá ser flexible; toda vez que, debido a las características propias del sector tecnológico que pretende intervenir, no sería posible anteponerse a todas las situaciones de alto riesgo que se generarán en el futuro (14). Así mismo, el capítulo dos del texto se ocupa de los requisitos con los que deben cumplir los sistemas de riesgos altos, entre los que sobresalen: (a) implementación de sistemas de administración del riesgo; (b) parámetros para el manejo de datos de entrenamiento y uso; (c) actualización y publicidad de la documentación técnica; (d) implementación de una bitácora de funcionamiento del sistema; (e) transparencia y entrega de información a usuarios; (f) supervisión humana durante la operación; y (g) niveles de certeza, robustos y ciberseguridad (46-53). Sumado a lo anterior, el capítulo 3 se ocupa de las obligaciones frente a los prestadores del servicio, usuarios y otros. Frente al apartado de la referencia, se observa que únicamente genera una obligación de supervisión por parte de las personas implicadas en la cadena de consumo (productor, importador, vendedor y usuario), frente a los requisitos anteriormente determinados en el capítulo 2.

Vale la pena resaltar que, de acuerdo con lo observado, la propuesta legislativa de la Comisión Europea basa la administración de los riesgos en un modelo de transparencia y "buenas prácticas" empresariales, generando la obligación de registro e informe ante las

[9] Por IIAA de riesgo alto se entienden: *"• Sistemas de IA destinados a ser utilizados como componentes de seguridad de productos sujetos a una evaluación previa de la conformidad por terceros; • otros sistemas autónomos de IA con implicaciones principalmente en los derechos fundamentales que se enumeran explícitamente en el Anexo III."* (Comisión Europea, 2021: 14).

autoridades, en cabeza de los sujetos anteriormente enunciados, como lo señala el capítulo 4. También resulta interesante, analizar como el título IV, numeral 3, genera una obligación directa a los usuarios de DF de informar que este tipo de tecnología fue utilizada en el proceso de generación de contenido (70).

Así mismo, es necesario resaltar que de acuerdo con la Comisión Europea (2021b) la implementación de esta regulación, a grandes rasgos, tiene como fin la creación de condiciones que permitan un desarrollo y uso confiable de IIAA en la UE, toda vez que identifican que esta tecnología está ligada a la optimización de beneficios sociales y al crecimiento económico, así como a la mejora de la competitividad global de la UE.

4.3.Consideraciones

A partir de lo analizado se evidencia una tendencia global hacia la regulación de las actividades relacionadas con las IIAA, por lo que la proyección jurídica de modelos de regulación y sus diferentes impactos, en lo relativo a las responsabilidades en las que podrían incurrir las personas sujetas a estas, se ha convertido en una necesidad central para las disciplinas jurídicas.

Vale la pena resaltar, de conformidad con lo analizado, que uno de los principales intereses por parte de los Estados que han empezado a legislar respecto a este tipo de tecnologías reside en mantener su competitividad a nivel mundial convirtiéndose en centro de innovación, desarrollo y comercio alrededor de las IIAA. En esa medida, es necesario prestar especial atención a la forma en que la nueva legislación termina desenvolviéndose, toda vez que la experiencia ha demostrado que ciertos agentes del mercado buscan incidir en los procesos legislativos a través del cabildeo, generando regulaciones favorables a sus intereses.

En esa medida, si bien la iniciativa de legislación está siendo tomada por los principales países "productores" de IIAA, resultaría fundamental que otros gobiernos inicien a prestar atención a estas tendencias. Sobre todo, cuando consideramos elementos como "soberanía informática". Se ha logrado identificar que, actualmente, el motor de la regulación es el evitamiento de capacidades de competencia. Por lo tanto, cada periodo legislativo que pase, sin que los países o asociaciones de países, que no tienen capacidades de negociación representativa ante la sociedad internacional, generen un avance hacía una regulación efectiva de estas nuevas tecnologías, será un paso hacia adelante con relación al ensanchamiento de la "brecha digital" anteriormente reseñada y, en consecuencia, hacía la pérdida de soberanía ante potencias como Chine y/o la UE.

5. Conclusiones y sugerencias

Después de analizar diferentes casos generados por el desbordamiento de riesgos, relacionados con el uso y/o desarrollo de inteligencias artificiales, resulta necesario comprender que este es un fenómeno multidimensional, por lo que la interdisciplinariedad será necesaria para hacerle frente al mismo. Igualmente, se evidencia, en el caso del acta de inteligencia artificial, una legislación de carácter extremadamente general, con obligaciones abstractas y poco especificas frente a la incidencia particular que los diferentes miembros de la cadena de producción y empleo pueden tener frente a la concreción de riesgos. Así mismo, se evidencia un enfoque extremadamente sesgado hacía las capacidades productivas y de optimización de las IIAA, lo que parece que la persona ha sido desplazada del foro de la regulación y se ha olvidado que finalmente serán los individuos aquellos que deberán sufrir los efectos adversos de un desarrollo y/o uso negligente o malintencionado de las IIAA.

Con base en lo anteriormente expuesto, a continuación, se proponen una serie de principios/parámetros que, se considera, deberían formar el núcleo y punto de partida de cualquier regulación y/o resolución de conflicto que gire alrededor del efecto dañino que las IIAA pueden tener sobre las personas[10]:

(a) **Principio de desarrollo *pro persona***: es necesario que los ordenamientos jurídicos se centren particularmente en que el desbordamiento de los riesgos recae sobre las personas. En esa medida, la legislación debe centrar al individuo, y su dignidad, como fundamento rector de cualquier legislación relativa a las IIAA. De esta manera, los Estados deberían enfocar sus políticas públicas, relacionadas con las IIAA, no desde un enfoque mercantilista e individualista; sino, más bien, desde una perspectiva de protección al humano. Por ejemplo, la implementación de tecnologías basadas en IIAA en los procesos relativos a la administración pública deberá estar acompañado por políticas públicas como la "alfabetización digital", con el fin de evitar fenómenos como el crecimiento de la brecha digital y otras formas de desigualdad.

Este principio sería transversal al resto de máximas propuestas, toda vez que pretende desarticular una lógica de regulación utilitarista que relacione el interés común con el desarrollo del sector tecnológico de la referencia. Derivado de esto, los siguientes principios descritos tendrán un enfoque a partir de los derechos e intereses de la persona expuesta a la IIAA. En esa medida, tanto las políticas publicas como la resolución de conflictos deberán

[10] Cabe aclarar que algunos de estos principios se encuentran presentes en la regulación publicada, sobre todo la europea. No obstante, se ha evidenciado un claro énfasis en la potencialización del mercado de las IIAA y poco énfasis en la protección de la personas. Frente a la regulación China, esta no ha sido publicada en otros idiomas, por lo que únicamente se logró trabajar con fuentes secundarias, a partir de las cuales no se evidencia un enfoque *pro persona* en la regulación.

desarrollarse alrededor del bienestar de los individuos humanos, por lo que esta sería tanto un principio rector legislativo como un regla de interpretación normativa.

(b) **Principio de especialidad de materia**: una regulación general y poco especifica no tendrá el impacto requerido para administrar las diferentes fuentes de riesgo identificadas. Por lo tanto, es necesario que, por una parte, la legislación se ocupe caso a caso de los riesgos específicos (utilizando macro-categorías como las identificadas en este trabajo) o que los proyectos de Ley consagren normativas tipo marco que puedan ser desarrolladas posteriormente y a la brevedad, por entidades del orden administrativo/ejecutivo. Así mismo, es necesario que las obligaciones reconocidas para cada actor relacionado en el uso/desarrollo de IIAA sean pensadas desde la especificidad de sus funciones, toda vez que esto generará mayor seguridad jurídica y confianza en el sistema. En esa medida se debe considerar la capacidad de injerencia que tendrían: programadores, científicos de datos y/o operarios, en los resultados generados por IIAA. Esto podría basarse en teorías basadas en el dominio del hecho y/o la posición de garante que los actores asumirían a lo largo del desarrollo y/o implementación de estos sistemas.

(c) **Principio de precaución**: este principio tendría como fin adelantarse a los principales daños analizados y sería la base para la generación de diferentes figuras como:

i. Prohibición de implementación de IIAA cuyo proceso de aprendizaje haya sido no supervisado en contextos relacionados con el ejercicio y goce de derechos fundamentales/humanos.

ii. Regulación de uso de IIAA en contextos específicos. Por ejemplo, en procesos democráticos, decisiones médicas, sistema justicia, sistema financiero y otros que incidan en los derechos fundamentales de las personas. Esto permitiría generar claridad alrededor

de cuales serían las conductas licitas que evitarían la generación de daños a los destinatarios de las conductas desplegadas por las IIAA.

iii. Obligación de fallo seguro de la IA. Lo que implicará que el sistema esté programado para que este opte por que los falsos positivos y los falsos negativos que pueda llegar a producir se den dentro de un espectro de seguridad, donde el error no represente un potencial daño irremediable.

iv. Obligación de control y registro. Como se ha evidenciado, algunas IIAA representan una amenaza debido a la naturaleza de sus capacidades; sin embargo, su prohibición representaría un retroceso tecnológico no justificado, por lo que el registro de estas tecnologías debería ser obligatorio por parte de los desarrolladores y usuarios que deseen relacionarse con estos sistemas. En esa medida, sería necesario que los Estados dispusieran de un órgano policivo ante el cual los particulares deberían acudir para registrar las tecnologías riesgosas y, así mismo, este órgano sería el encargado de fiscalizar el uso seguro de los sistemas.

(d) **Principio de actualización:** se ha evidenciado que, si bien los estados han buscado generar herramientas de protección informática, relativa al manejo de datos de los usuarios, para evitar fuentes de discriminación, el nivel actual de la tecnología ha dejado dichos cuerpos normativos obsoletos. Por lo tanto, es necesario que la legislación relativa a IIAA esté encaminada a identificar y transformar el sistema jurídico del que hace parte, de manera integral. Así mismo, debe prever la necesidad de actualización constante, tal como se explicó en el principio de especialidad. Es así, como el derecho administrativo, y en particular los decretos ley, serían de gran valía. Este tipo de herramientas permiten un dinamismo contaste, e igualmente controlado, entre la realidad del sector tecnológica, sus nuevos productos y la

regulación estatal. En esa medida, sería necesario también que los Gobiernos conformaran cuerpos administrativos, como los reseñados en EE. UU., cuyas funciones se encaminaran a estudiar constantemente las innovaciones tecnológicas y proyectaran, periódicamente[11], informes de actualización legislativa.

(e) **Principio de transparencia material:** este principio busca obligar a que los desarrolladores/implementadores de inteligencias artificiales sean precisos y, sobre todo, claros frente a las capacidades, limitaciones, funciones, herramientas, modo de uso, etc., de sus tecnologías. Debe hacerse hincapié en que la materialidad de este principio se basa en la posibilidad de que cualquier persona sea capaz de comprender el mensaje transmitido, evitando la hiper-complejización de la información. Por ejemplo, con el boom de tecnologías de procesamiento de texto como Chat-GPT o Auto-GPT es necesario que los usuarios comprendan tanto sus límites y capacidades de manera holística. Lo anterior, tendría como fin generar una mejor experiencia para los usuarios. Esto permitiría tomar decisiones informadas por parte de las personas que deban exponerse ante la IIAA y también permitiría una mejor operabilidad por parte de los no expertos.

Es importante resaltar que la propuesta anteriormente planteada podría ser aplicada tanto a las regulaciones nacionales y supranacionales que nazcan con el pasó del tiempo. Esto se cimienta en que las diferentes obligaciones, deberes y derechos, anteriormente identificados, se encuentran en cabeza de individuos y de Estados. Por lo tanto, la obligaciones, en cabeza de los Estados, de protección, respeto y garantía de derechos humanos serían exigibles a los países a partir los efectos que las IIAA tienen sobre las personas bajo su jurisdicción. Ahora

[11] Se considera que la periodicidad debería darse de acuerdo a las necesidades del momento, no obstante, la proyección de 3 o 4 resoluciones anuales parecería ser suficiente.

bien, frente a los sistemas internos es necesario indicar que las herramientas de cumplimiento (coercitivas) existentes en los Estados para hacer valer estos principios partirían desde la implementación de sanciones administrativas o penales. A partir de lo investigado se considera que las medidas administrativas serían menos intrusivas, sobre todo tomando en cuenta la naturaleza de *ultima ratio* del derecho penal. Por lo tanto, el órgano policivo encargado reseñado anteriormente, no solo sería un órgano de control frente a la inscripción de nuevas tecnologías, sino que dentro de sus capacidades fiscalizadores se encontrarían las de sanción. Adicionalmente, es innegable que, en algunos casos, el daño generado a partir de del uso/desarrollo de IIAA atentará de manera relevante contra ciertos bienes jurídicos tutelados por el derecho penal y, por lo tanto, sería posible imponer sanciones del tipo penal.

Finalmente, es claro que la regulación en materia de IIAA será una realidad; no obstante, es necesario que la reflexión académica desarrollada previa a la entrada en vigor de esta o posterior a esta, esté encaminada a aclarar las dudas relativas al funcionamiento de lógicas jurídicas pensadas para un sistema que operaba a una velocidad distinta, donde la autonomía era un concepto que únicamente podía relacionarse con las personas y donde la injerencia de una u otra persona en un curso causal no dependía de datos supremamente volátiles y muy poco visibles. Por lo tanto, la meditación sobre adaptación de las categorías clásicas jurídicas frente al fenómeno de las IIAA más que un capricho académico es una necesidad.

6. Lista de referencias bibliográficas

Doctrina

Arévalo, J., Bayona, R., & Dewar, R. (2015). "*El problema de la brecha tecnológica: Un asunto de cultura.*" Revista Sinapsis,

Barocas, S., & Selbst, A. D. (2016). "*Big Data's Disparate Impact. California Law Review, 104(3),*" Recuperado de: https://doi.org/10.2139/ssrn.2477899

Beck, U. (1998). "*Teoría de la sociedad del riesgo: Las consecuencias perversas de la modernidad.*" Paidós.

Beck, U. (2006). "*La sociedad del riesgo.*" Editorial Iberica

Blocher, J. (2021). "*When guns threaten the public sphere: a new account of public safety regulation under heller.*" The Yale Law Journal, N. 131(1), Pg.78-155.

Candelon, F., Garrigos-Simon, F. J., & Evrard-Todeschi, N. (2021). "*AI regulation is coming.*" Harvard Business Review, Ed.99(2), Pg.126-135.

Chadha, A., Kumar, A., Jain, A., & Goel, P. (2020). "*Deepfake: An Overview. In Proceedings of Second International Conference on Computing Communications, and Cyber-Security*" (pp. 463-468). Springer, Singapore.

Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). "*Artificial intelligence, bias and clinical safety.*" BMJ Quality & Safety, Ed. 28(3), Pg. 231-237.

Cohen, G., & Méndez, R. (2012). "*La sociedad del riesgo: amenaza y promesa*" (1ra ed.). Paidós. p. 27.

Cook, P. J. (1995). "*Regulating Gun Markets.*" Oxford University Press.

Crovi, D. M. (2010). "*Jóvenes, migraciones digitales y brecha tecnológica. En C. Sigal, E. Livchits, & M. Mayer (Eds.), Entre la brecha digital y la participación ciudadana: nuevas formas de exclusión e inclusión en la sociedad de la información*" (pp. 37-68). Editorial Paidós.

Dubber, M., Pasquale, F., & Mittelstadt, B. (Eds.). (2020). "*The Oxford Handbook of Ethics of AI.*" Ed.Oxford University Press.

Foucault, M. (1966). "*Las palabras y las cosas.*" México D.F.: Siglo XXI Editores.

Giddens, A. (2000). "*La Tercera vía: La renovación de la socialdemocracia [The Third Way: The Renewal of Social Democracy].*" Madrid, España: Editorial Taurus

Jakobs, G. (1996). "*La imputación al tipo objetivo.*" Editorial Civitas.

Lombana Bermudez, A. (2018). "*La evolución de las brechas digitales y el auge de la Inteligencia Artificial (IA).*" Revista Mexicana De Bachillerato a Distancia, Recuperado de: https://doi.org/10.22201/cuaed.20074751e.2018.20.65884

McKinsey Global Institute (MGI). (2017). "*Artificial intelligence: The next digital frontier? [Discussion paper].*" McKinsey & Company. Recuperado de: https://www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx

Nelson GS. "*Bias in Artificial Intelligence.*" N C Med J. 2019 Jul-Aug;80(4):220-222. doi: 10.18043/ncm.80.4.220. PMID: 31278182.

Nietzsche, F. (2000). El Nacimiento de la Tragedia. Ed. Penguin Classics.

Nyéléni. (2015). "*Declaración de Nyéléni sobre la Soberanía Alimentaria.*" Recuperado de https://nyeleni.org/IMG/pdf/DeclNyeleni-es.pdf

Ziad Obermeyer et al. ,Dissecting racial bias in an algorithm used to manage the health of populations. Science366, Recuperado de: https://www.science.org/doi/10.1126/sciadv.add2315

Rousseau, J.J. (1985).”*El Contrato Social*”, Editorial Madrid: Alba.

Romero, A. (2020). “*La brecha digital generacional.*” Editorial: Universidad De Vigo.

Siegel, R. B. & Blocher, J. (2020). “*Why regulate guns.*” Ed. Oxford University Press.

Silber, J. & Manyika, J. (2018).”*Notes from the AI frontier: Tackling bias in AI (and in humans).*” McKinsey & Company. Recuperado de: https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/mgi-tackling-bias-in-ai-june-2019.ashx

Stone, P., et al. (2016). “*Artificial intelligence and life in 2030. One hundred year study on artificial intelligence: Report of the 2015-2016 study panel,*” Stanford University, Ed. Stanford. University. Recuperado de https://ai100.stanford.edu/2016-report

Van Huijsteee, M., Costello, C., Burgess, M., & Llobet Rodriguez, P. (2021). “*Tackling deepfakes in European policy.*” European Parliament Policy Department for Citizens' Rights and Constitutional Affairs. Recuperado de https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf

Wester, M. (2020). “*The Emergence of Deepfake Technology: A Review.*” Technology Innovation Management Review. Recuperado de: https://timreview.ca/sites/default/files/article_PDF/TIMReview_November2019%20-%20D%20-%20Final.pdf.

Casos

Brief of Defendant-Appellant. Loomis v. Wisconsin, 881 F.3d 437 (7th Cir. 2018).

Corte Constitucional. (1995). Sentencia C-296 de 1995. Recuperado de https://www.corteconstitucional.gov.co/relatoria/1995/c-296-95.htm

Harvard Law Review. (2017). Recent Cases, 130(5), 1534-1536. Recuperado de https://harvardlawreview.org/archives/vol-130-no-5/

Informes, Noticias y Otros

Abc News. (2006). "*Does the soda you drink reveal how you vote?*" Recuperado de https://abcnews.go.com/GMA/story?id=2623263&page=1

Banco Mundial (2021). "*Usuarios de internet (% de la población)*". Recuperado de https://data.worldbank.org/indicator/IT.NET.USER.ZS

Consejo Nacional de Ciencia y Tecnología de la Casa Blanca. (2016). "*Preparing for the future of artificial intelligence.*" Recuperado de https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

ChatGPT. (2023). "Hablo español". *OpenIA.* Recuperado de: https://chat.openai.com/c/115171da-d458-49c5-b584-fec187a98852

Cui, X., & Chen, Y. (2022). "*What China's Algorithm Registry Reveals About AI Governance*." Carnegie Endowment for International Peace. Recuperado de: https://carnegieendowment.org/2022/12/09/what-china-s-algorithm-registry-reveals-about-ai-governance-pub-88606

Destatis. (2022).*"Bildungsabschlüsse in Deutschland [Educational qualifications in Germany]*." Destatis Recuperado De https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsstand/Tabellen/bildungsabschluss.html

European Commission. (2021). "*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.*" Official Journal of the European Union, Recuperado de: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

European Commission. (2021). "*Executive summary of the impact assessment report.*" Recuperado de https://ec.europa.eu/info/sites/default/files/impact-assessment-artificial-intelligence-regulation_en.pdf

European Commission. (2021c). "*Regulatory framework for artificial intelligence.*" Recuperado de: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Harwell, D. (2019). "*Top AI researchers race to detect 'deepfake' videos: 'We are outgunned"*". The Washington Post Recuperado de: . https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/

McAuley, E. (2020). "*What happened with Cambridge Analytica?*", BBC News. Recuperado de https://www.bbc.com/news/technology-43465968

New York State Department of Financial Services. (2021). Report on Apple Card Investigation. Recuperado de: https://www.dfs.ny.gov/system/files/documents/2021/03/rpt_202103_apple_card_investigation.pdf

OECD. (2021). "*AI Policy Observatory - Policy Initiatives.*", OCDE. Recuperado de: https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-24274

O'Sullivan, L. (2021). "*How the law got it wrong with Apple Card.*" TechCrunch. Recuperado de: https://techcrunch.com/2021/08/14/how-the-law-got-it-wrong-with-apple-card/

Pradhan, P. (2020). "*AI Deepfake. The Goose is cooked?*" University of Illinois Law Review. Recuperado de: https://illinoislawreview.org/blog/ai-deepfakes/

ProPublica. (2016). "*How We Analyzed the COMPAS Recidivism Algorithm*". ProPublica. Recuperado de: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

ProPublica. (2016). "*Machine Bias.*" ProPublica. Recuperado de https://www.propublica.org/documents/item/3063040-Machine-Bias-Report.pdf].

Statista. (2021). "*Education gender gap worldwide by level.*" Recuperado de: https://www.statista.com/statistics/1212278/education-gender-gap-worldwide-by-level/

The New York Times. (2018). "*Cambridge Analytica and Facebook: The scandal and the fallout so far.*" Recuperado de: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

United Nations. (2021). "*AI Ethics body launched by UN chief to tackle impact of rapid technological change.*" Recuperado de: https://news.un.org/en/story/2021/11/1106612

Vigdor, N. (2019). "*Apple Card investigated after gender discrimination complaints.*" The New York Times. Recuperado de: https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html