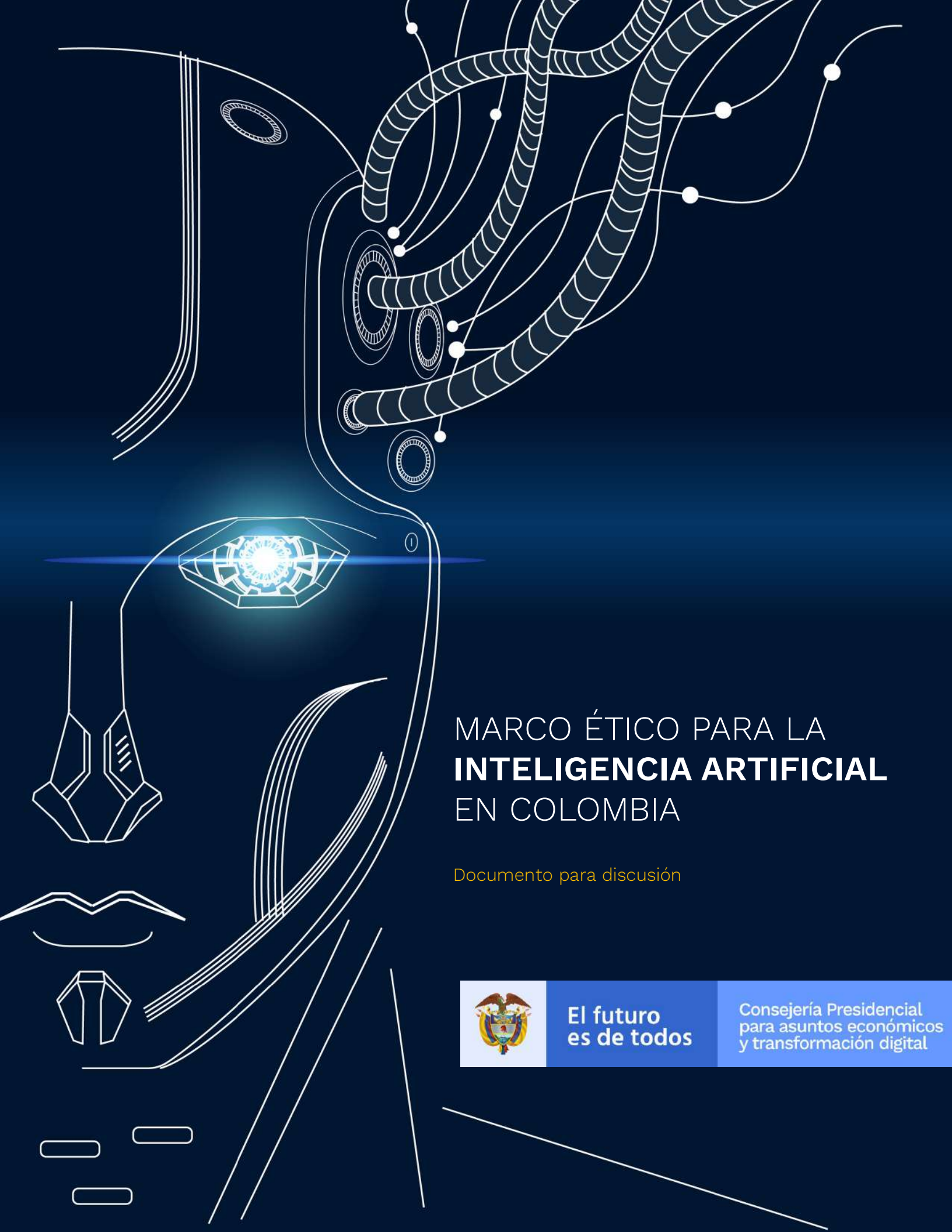


MARCO ÉTICO PARA LA **INTELIGENCIA ARTIFICIAL** EN COLOMBIA

Documento para discusión

El futuro es de todos

Consejería Presidencial para asuntos económicos y transformación digital

MARCO ÉTICO PARA LA **INTELIGENCIA ARTIFICIAL** EN COLOMBIA

Documento para discusión

Autor:
Armando Guío Español
Consultor de la Corporación Andina de Fomento

Fecha:
Agosto, 2020

Entidades involucradas en la realización del documento

Reconocimientos

La vocación de este documento es la de servir como un insumo a la conversación nacional, urgente y necesaria, sobre el marco ético del desarrollo de la inteligencia artificial en Colombia. En noviembre 2019, el Gobierno de Colombia adoptó una Política Nacional para la Transformación Digital e Inteligencia Artificial contenida en el documento CONPES 3975. Este documento busca por lo tanto nutrir el diálogo nacional sobre el desarrollo de la inteligencia artificial en el marco de la recuperación. En este sentido, no representa las opiniones oficiales ni del Gobierno de Colombia, ni de CAF Banco de Desarrollo de América Latina.

Este documento fue preparado por Armando Guío Español, afiliado del Berkman Klein Center for Internet & Society de la Universidad de Harvard. Fue comisionado por la Dirección de Innovación Digital del Estado de CAF Banco de Desarrollo de América Latina para la Consejería para Asuntos Económicos y Transformación Digital de la Presidencia de la República de Colombia. Se enmarca en la agenda de CAF sobre el uso responsable de la inteligencia artificial en el sector público en América Latina. Ha sido supervisado por María Isabel Mejía con el apoyo de Nathalie Gerbasie y revisado por María Isabel Mejía, Víctor Manuel Muñoz y Carlos Santiso.

CONTENIDO

La discusión ética debe partir de un entendimiento de los principios y modelos que han sido propuestos a nivel mundial, su impacto en distintas etapas de desarrollo de la inteligencia artificial, aquellos que podrían aplicar al caso colombiano y los mecanismos de implementación existentes.

I
INTRODUCCIÓN

Una de las tareas fundamentales que ha dispuesto el Gobierno colombiano es la de tener un marco ético propio para el desarrollo de la Inteligencia Artificial en el país que sea aplicable tanto al sector público como privado. Esto implica el diseño de un marco transversal que pueda ser aplicable a distintos actores, considerando diversidad de intereses y opiniones.

La discusión alrededor de la ética en la Inteligencia artificial no es nuevo y se ha venido desarrollando durante varias décadas. Desde mitades del siglo XX e incluso mucho antes, la filosofía se ha encargado de estudiar este tema a profundidad. Por lo tanto, no se debe desconocer la historia que rodea esta temática y los distintos esfuerzos que se han generado por generar un marco comprensible alrededor del desarrollo de una tecnología que pone a prueba nuestro entendimiento de la realidad.

Es por esto que surgen varios elementos a considerar en la construcción de un marco ético inteligencia artificial en Colombia:

1 Justificación de este marco

Una de las primeras preguntas que puede surgir al momento de diseñar un marco ético a la inteligencia artificial en Colombia es por qué se debe pensar en un proyecto de estas características en este país. Existen dos elementos fundamentales que hacen de éste un ejercicio necesario y prioritario, más allá de que esto sea una tendencial global que varios países vienen proponiendo.

En primer lugar, los efectos que puede tener la implementación de la inteligencia artificial en el país son variados y hay distintas implicaciones éticas de la forma como esta tecnología llegue a ser utilizada. La experiencia internacional y la evidencia recolectada a nivel mundial demuestran que los sistemas de inteligencia artificial utilizados en diversos sectores como la justicia, la actividad policial y en el mercado financiero, entre otros casos, pueden llevar a las prácticas discriminatorias, injustas y con implicaciones sociales no deseables. Asimismo, el uso de esta tecnología puede entrar en conflicto con varios derechos fundamentales y derechos humanos que Colombia se ha comprometido a respetar y defender desde la Constitución Nacional de 1991 y los tratados internacionales que ha ratificado. Es por esto que los límites éticos definidos resultan esenciales en el despliegue de esta tecnología en el país.

A esto se une la coyuntura actual provocada por el COVID-19 y que hace que la automatización y el uso de sistemas de Inteligencia artificial en diversos sectores pueda haberse acelerado en detrimento de la mano de obra de los trabajadores

colombianos. Esto lleva necesariamente a que el país deba abordar una discusión sobre la forma como miles de trabajos pueden ser reemplazados, el lugar de los seres humanos en la cuarta revolución industrial y en últimas el rol del ser humano en la implementación de estos sistemas. Esto hace parte de discusiones propias del rol del ser humano en la implementación de esta tecnología y en el grado de control que tienen sobre sus decisiones y acciones.

En segundo lugar, la ética de la Inteligencia artificial resulta ser esencial como un primer paso hacia la creación de límites sociales deseables en el uso de esta tecnología. La discusión ética brinda un punto de partida y permite llegar a una serie de consensos sociales sobre el uso de esta innovación, y que tienen efectos menos disruptivos en la innovación, a diferencia de un marco regulatorio. Por lo tanto, permite un mayor campo de exploración de esta tecnología, su desarrollo futuro y sus efectos, que permitan consolidar posteriormente una evidencia que sirva para justificar medidas regulatorias o de otra índole.

De esta forma la discusión ética permite abordar esta tecnología desde una posición de discusión y de experimentación en el diseño de medidas específicas materializar esos principios éticos. No obstante, no se debe subestimar el profundo impacto que tiene la discusión ética, ya que brinda los elementos fundamentales sobre los cuales seguramente se procederá más adelante a diseñar una regulación sobre el tema. Es por esto que esta discusión ética es al final una discusión con implicaciones regulatorias, pues busca analizar lo que serán las bases filosóficas y los principios iniciales que guían el uso de la Inteligencia artificial en el país y la normativa que se desarrolle al respecto.

2 Estado de desarrollo tecnológico

Las discusiones sobre el marco ético de la inteligencia artificial aplicable a Colombia deben centrarse en el desarrollo actual de esta tecnología, entendiendo sus verdaderas funcionalidades y los distintos tipos de inteligencia artificial que se han venido desarrollando y que están siendo aplicados. En la actualidad, contamos con sistemas de inteligencia artificial específicos y limitados a cierto tipo de procesamiento y tareas. Es así como hay sistemas de Machine Learning y Deep learning, como dos ejemplos del tipo de Inteligencia artificial que se encuentran disponibles. Es sobre este estado de desarrollo que debe recaer una discusión actual de la ética de la inteligencia artificial y es así como lo han abordado distintas autoridades alrededor del mundo.

Establecer discusiones éticas sobre si los sistemas de inteligencia artificial tienen la capacidad de seres conscientes y de nuestra convivencia con este tipo de seres desconoce que en la actualidad no hemos llegado a un estado de desarrollo que responde más a la ficción que a una realidad. Lo que la literatura ha considerado como una inteligencia artificial general no es propio de la tecnología actual que encontramos disponible. Llevar la discusión sobre este tipo de inteligencia artificial no solo resulta desgastante, sino que puede desviarnos de los puntos que en la actualidad son más apremiantes, como el procesamiento de datos o los modelos que guían el diseño actual de un algoritmo y que afectan a miles de personas hoy en día.

3 Las distintas etapas de inteligencia artificial

El diseño de un marco ético de estas características debe considerar la existencia de distintas etapas que describen integralmente la elaboración y despliegue de un sistema de inteligencia artificial (*algorithmic chains*). Dentro de esta 'cadena encontramos' cuatro fases fundamentales:

1. **Diseño:**
   Esta fase incluye tanto el diseño de tecnologías basadas en IA (e.g. cómo se recolectan los datos, la audiencia a la que se dirige una herramienta de IA) así como el diseño de sistemas que gobiernan la IA y las implicaciones que se deben considerar antes de pasar a la siguiente fase. (AI & Inclusion Staff, s.f.).

2. **Desarrollo:**
   La fase de desarrollo es la fase de producción de un sistema autónomo que sigue al proceso de diseño. Preguntas en la categoría de desarrollo pertenecen a la incorporación de herramientas y métodos a las tecnologías basadas en IA, marcos para el desarrollo de IA y cuáles herramientas de IA se desarrollan para quién. (AI & Inclusion Staff, s.f.).

3. **Implementación:**
   La etapa de implementación incluye la distribución, el uso, la ubicuidad y la ejecución de tecnologías basadas en IA dentro de la sociedad en múltiples niveles, incluyendo los ecosistemas locales, nacionales y globales. (AI & Inclusion Staff, s.f.).

4. **Evaluación e Impacto:**
   La fase de evaluación / impacto incluye la medida y el entendimiento del impacto de tecnologías de IA, incluyendo maneras para evaluar los efectos de sistemas autónomos en diferentes agentes dentro de la sociedad. (AI & Inclusion Staff, s.f.).

Al momento de abordar la propuesta ética de este marco y los principios propuestos se podrá hacer referencia a cualquiera de estas fases, lo cual significa que algunas implicaciones se circunscriben a algunas de estas etapas específicas.

4 Insumos que provee el escenario internacional

La ética en la inteligencia artificial se ha venido desarrollando alrededor del diseño, desarrollo e implementación de esta innovación, especialmente de *machine learning* y *deep learning*. Por consiguiente, resulta deseable que el gobierno colombiano también centre su labor en este tipo de tecnologías y en los principios aplicables a estos sistemas. Sin embargo, el Gobierno colombiano no tiene por qué partir de cero en esta tarea. Esto no solo resultaría ser ineficiente, sino que es innecesario y desconoce que desde hace ya algunos años se vienen proponiendo a nivel mundial una serie de principios que guían el desarrollo de esta tecnología. Entidades privadas, gobiernos, ONGs y organismos multilaterales ya vienen haciendo propuestas al respecto. Incluso dentro de varias de las políticas y estrategias nacionales de inteligencia artificial que se han propuesto desde el año 2017, varios países ya se han atrevido a formular una serie de principios.

Esto no significa que el Gobierno colombiano simplemente deba hacer un proceso de selección y aplicación de los principios que considere más relevantes. El verdadero reto está en seleccionar aquellos que respondan a las necesidades del país, dotarlos de sentido y contenido, establecer las implicaciones que tendrían para el contexto propio del país y la forma como se verían materializados en los sectores públicos y privados de Colombia.

Por lo tanto, la propuesta es que el Gobierno colombiano conozca los principios conocidos a nivel internacional y dirija todos sus esfuerzos a establecer cuales de estos aplicarán en el país, como deben ser entendidos en las distintas etapas de desarrollo de la inteligencia artificial y diseñe mecanismos específicos para su implementación y materialización. Esto no impide que Colombia pueda generar principios propios y tal vez nunca antes vistos en el escenario internacional. Sin embargo, antes de generar este tipo de propuesta es importante que haya un conocimiento extenso de los insumos que ya provee el sistema internacional y que justifique la necesidad de proponer alguno realmente nuevo. Además, esto mismo ya se señalaba dentro del Conpes 3975 de 2019 que, como hoja de ruta para el diseño de este marco ético, establecía una serie de principios que deben ser tenidos en cuenta y la necesidad de valerse de la experiencia internacional alrededor del tema.

5 Una ética transversal

Dado que el Gobierno colombiano quiere que el marco ético propuesto aplique al sector público y privado, es necesario considerar las realidades y particularidades propias de cada sector. Es entendible que el Gobierno busque desarrollar un marco general, dado un criterio de igualdad y por el deseo de que no existan asimetrías en la forma como cada uno de estos sectores aborda este tipo de tecnología. Se espera que este marco sirva como un insumo general que permita después a cada entidad dentro de alguno de estos sectores desarrollar su propio marco ético que siga una serie de particularidades y mecanismos propios de implementación.

Ahora bien, la principal consecuencia de tener un marco transversal es que hace necesario que se parta de grandes principios que pueden ser aplicados a la sociedad en general. Así mismo, señalará una serie de mecanismos implementar dichos principios sin entrar en especificidades en cuanto a las entidades y personas encargadas de este proceso o las metodologías propias a seguir en cada sector, dado que esto dependerá de la lógica propia de cada entidad.

6 Una conversación nacional

Dado todo lo anterior, las implicaciones de esta discusión ética son variadas y profundas para la sociedad colombiana. A través de este esfuerzo se están definiendo lineamientos que impactarán la forma como una generación de colombianos abordará la innovación más importante de los últimos siglos que haya conocido la humanidad. Dichos principios tendrán un alto impacto en distintas políticas públicas, normativas y otros documentos oficiales al igual que guiará la innovación dentro de la Cuarta Revolución Industrial. Tal discusión se convertirá en la base de una política de Estado y por ende requiere de una discusión previa con distintos sectores de la sociedad que analicen los principios que se proponen desde el Gobierno y sus implicaciones.

Es por esto que la construcción final de este marco debe ser un ejercicio democrático e incluyente que involucre a distintos actores en el ámbito nacional e internacional. **La vocación de este documento es la de servir como un insumo inicial que sirva de punto partida para que dicha discusión tenga lugar de forma organizada y con criterios claros que guíen esta conversación.** Un debate que no tenga una hoja de ruta con unos elementos de discusión concretos difícilmente generará un producto preciso. Por esto, este insumo se hace necesario, ya que además da mayor desarrollo a los temas éticos ya mencionados

en el documento Conpes 3975 de 2019, y que ya han sido resaltados por entidades internacionales como la Organización para la Cooperación y Desarrollo Económicos (OCDE) o el Banco Interamericano de Desarrollo (BID). Esta última entidad incluso señaló en su reporte de 2020:

> “El Government AI Readiness Index de 2019, producido con el apoyo de Oxford Insights y el Centro Internacional de Investigaciones para el Desarrollo (IDRC por sus siglas en inglés), muestra que los países de la región se enfrentan a tres desafíos cuando se trata de aprovechar el uso de la IA en favor del bien común: políticas, capacidad y recursos adecuados. En primer lugar, hasta la fecha ALC no cuenta con un enfoque coherente de política y tampoco con estándares éticos definidos. México, Colombia, Uruguay y Argentina están fijando actualmente políticas y estrategias de lA. Colombia, por ejemplo, por medio del documento CONPES 3975 definió su Política Nacional para la Transformación Digital e Inteligencia Artificial. Allí se identifican lineamientos concretos que, a través de su implementación, generarán un marco coherente de política para el desarrollo ético y responsable de la IA”
>
> (Cabrol, González, Pombo, & Sanchez, 2020).

II

¿QUÉ ES LA ÉTICA DE LA INTELIGENCIA ARTIFICIAL?

Se considera La ética de la inteligencia artificial como una rama de la ética que analiza y evalúa los dilemas morales que se derivan del despliegue de esta tecnología en la sociedad. Desde hace siglos la idea de seres conscientes y pensantes ha sido utilizada por la filosofía para analizar características propias de los seres humanos y si estas son trasladables a otros objetos. De esta forma, se desarrollando conceptos como los de la consciencia y su importancia en esta rama, teniendo en cuenta que a través de nuestra conciencia individual tomamos conocimiento de nuestros principios morales más arraigados, nos motivamos a actuar conforme y evaluamos nuestro carácter, nuestro comportamiento y en últimas a nosotros mismos, conforme a estos principios (Giubilini, 2016).

En los últimos años se ha derivado una nueva aproximación a la ética de la inteligencia artificial, basada en la ética de los datos (*Data Ethics*). Esta ética se centra en el uso y analítica de los datos y los distintos sistemas e innovaciones que interactúan con esta información. Dada la importancia que la analítica de datos y el auge de *big data* ha tenido en los últimos años, esta es una de las éticas que mayor predominancia ha tenido y muchas de las propuestas éticas y de principios se derivan de este estudio.

Como lo afirman Floridi y Taddeo (2016), esta ética se basa en el fundamento proveído por la ética de la computación e informática, pero, al mismo tiempo, perfecciona la perspectiva apoyada hasta el momento en este campo de investigación, al cambiar el nivel de abstracción de las preguntas éticas, de ser centrado en la información a estar centrado en los datos.

Como lo señala el mismo Floridi, en este caso tenemos una transformación en la perspectiva debido a un cambio en los niveles de abstracción (LoA). Es decir que centrarnos de una definición más abstracta y general sobre la información y los diversos dilemas éticos que se pueden derivar de estos conceptos, se pasa a un enfoque centrado en los datos, especialmente aquellos que son utilizados para el desarrollo e implementación de sistemas de inteligencia artificial (Floridi & Taddeo, 2016).

De esta forma, se propone una definición de la ética de los datos que a su vez divide esta disciplina en tres criterios de análisis: una nueva rama de la ética que estudia y evalúa los problemas morales relacionados con los **datos** (incluyendo su generación, registro, adaptación, tratamiento, divulgación, diseminación y uso), **algoritmos** (incluyendo IA, agentes artificiales, *machine learning* y robots) y **prácticas** correspondientes (incluyendo innovación responsable, programación, *hacking* y códigos profesionales), con el fin de formular y apoyar soluciones moralmente buenas (e.g. códigos adecuados o valores correctos) (Floridi & Taddeo, 2016).

Esto también hace que el centro del estudio ético deba ser los sistemas y operaciones computacionales en los que estos datos son utilizados, más que en la variedad de tecnologías digitales que los facilitan (Floridi & Taddeo, 2016). Esta ética es valiosa dado que permite desarrollar buenas prácticas y conductas consideradas como moralmente buenas para abordar los dilemas éticos planteados por la colección y análisis de grandes bases de datos y sobre dilemas que van desde el uso de *big data* en investigación biomédica y en las ciencias sociales, al perfilamiento, publicidad y filantropía de datos, así como *open data* (Floridi & Taddeo, 2016).

Dado su sentido práctico y que permite discutir juicios éticos concretos y frente a los desarrollos tecnológicos actuales, esta es la ética en la que debe centrarse la propuesta de un marco transversal por parte del Gobierno colombiano. **La ética de los datos será analizada a lo largo de este documento como la ética aplicable a la inteligencia artificial que se diseña, implementa y desarrolla en Colombia.**

Así mismo, se tendrá en cuenta la forma como se divide esta ética al momento de analizar cada uno de los principios éticos propuestos y determinar cómo se relacionan con la ética de los datos, los algoritmos y las prácticas. Esto permitirá desarrollar una mirada integral de cada uno de los principios propuestos.

III

PRINCIPALES RETOS ÉTICOS Y EFECTOS NO DESEADOS DE LA ÉTICA DE LA IA

Como ya se ha señalado en este documento, **la implementación de la inteligencia artificial ya ha supuesto una serie de retos éticos, al igual que ha generado inquietudes alrededor del impacto que puede tener en algunos casos.**

Tal y como lo ha señalado el gobierno francés (CNIL, 2018), existen los siguientes retos específicos que se deben abordar: (i) posibles amenazas a la libertad de voluntad y responsabilidad; (ii) sesgos, discriminación y exclusión; (iii) perfilamiento algorítmico: personalización versus beneficios colectivos; (iv) buscar un nuevo balance al prevenir bases de datos masivas mientras se aumenta la IA; (v) calidad, cantidad y relevancia: el reto de los datos adaptados para IA, y (vi) la identidad humana frente al reto de la inteligencia artificial.

Según la Web Foundation, los principales efectos negativos que puede tener un sistema de inteligencia artificial se dividen en daños y discriminación (*Algorithmic Harm and Algorithmic Discrimination*). En cuanto a los daños algorítmicos, la Web Foundation (World Wide Web Foundation, 2017) ha explicado que los valores de cada sociedad subyacen la definición de daños. Al definir lo que un algoritmo no debe hacer (daño), emergen límites sólidos para lo que debe ser una función de optimización de un algoritmo (objetivos amplios). Asegurarse de que los algoritmos sean compatibles con la diversidad de valores que existe alrededor del mundo es ciertamente un reto. ¿Quién debe definir y determinar si los algoritmos han producido un daño? ¿En qué casos debemos promover que quienes hayan podido ser afectados por un algoritmo se integren al proceso de diseño?. La Web Foundation toma la definición de daño de la esfera legal, que define el daño como retrocesos que además se considera que están mal. Los riesgos antes mencionados se tratan de abordar desde el concepto de legitimidad que han propuesto distintas entidades internacionales, tales como la Web Foundation.

Por su parte, la discriminación puede ocurrir de dos formas (World Wide Web Foundation, 2017). Dos personas pueden ser iguales en aspectos relevantes, pero ser tratadas diferente, o las diferencias relevantes entre ellas no son reconocidas o tenidas en cuenta y las dos personas son tratadas igual. En el segundo escenario, al no tener en cuenta estos detalles relevantes, el resultado es injusto y en consecuencia. De esta manera, una persona puede esperar razonablemente un resultado que es injustamente impedido por un algoritmo, constituyendo un daño. Los países con altos y bajos ingresos enfrentan las mismas categorías de daños y amenazas provenientes de la toma de decisiones algorítmica.

Ahora bien, el impacto de estos daños puede ser ampliamente diferente, dependiendo de las protecciones legales existentes y los mecanismos de *accountability* (responsabilidad demostrada) implementados, especialmente para los grupos marginalizados (World Wide Web Foundation, 2017). En algunos países, la discriminación algorítmica y las predicciones inexactas pueden resultar en publicidad no deseada o en otros inconvenientes en las experiencias de los

consumidores, pero para los grupos marginados en contextos frágiles, se argumenta que la discriminación algorítmica puede llevar a agresiones sin control e incluso a exclusiones fatales de servicios y recursos públicos.

Como se puede observar, la mayoría de las preocupaciones está alrededor de la discriminación que puedan generar estos sistemas, como pueden profundizar las desigualdades y la posibilidad de tener sistemas que tomen decisiones automáticamente, sin que exista control y orientados por los prejuicios y la discriminación que ha sido establecida en su diseño o en la data que utiliza.

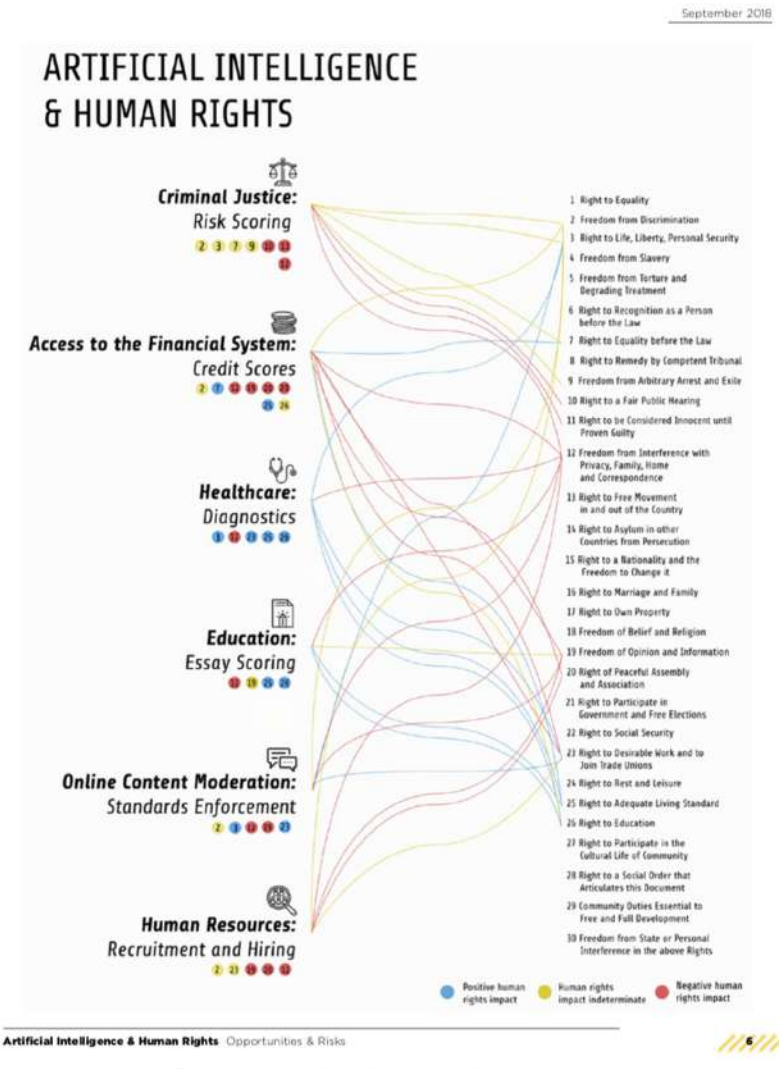
Ejemplo de esto han sido los **sistemas de reconocimiento facial** que se valen de esta tecnología. Existe una amplia evidencia de que esta tecnología causa daños graves, con frecuencia a personas de color y gente pobre (Crawford, y otros, 2019). En consecuencia, debe haber una suspensión en todos los usos de reconocimiento facial en los dominios públicos y sociales sensibles, donde este tipo de reconocimiento tiene riesgos y consecuencias que no se pueden remediar retroactivamente. Los legisladores deben suplementar esta suspensión con (1) requisitos de transparencia que permita a investigadores, hacedores de política pública y comunidades asesorar y entender la mejor forma de restringir y regular el reconocimiento facial, y (2) protecciones que brinden a las comunidades sobre las que se utiliza esta tecnología con el poder de tomar sus propias evaluaciones y rechazar la implementación.

Estos riesgos no solo pueden terminar afectando a la sociedad general, sino los derechos humanos fundamentales y propios de cada individuo. A través de una investigación de casos de estudio, miembros del Berkman Klein Center de la Universidad de Harvard definieron que los sistemas de inteligencia artificial que habían sido implementados en los Estados en distintos sectores habían tenido un impacto en distintos derechos humanos, como la libertad de expresión, el debido proceso, derecho de asociación e igualdad, entre otros.

A partir de su investigación lograr determinar los siguientes impactos específicos de esta tecnología en varios de estos derechos:

- **Sistemas de Inteligencia Artificial en el sistema penal para determinar nivel de riesgo de un individuo y grado de reincidencia:** conforme al estudio estos tienen un impacto negativo en el derecho a una audiencia pública justa, el derecho a ser considerado inocente hasta que se pruebe lo contrario y la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia.
- **Sistemas de Inteligencia Artificial en el sistema financiero para determinar el riesgo crediticio de un individuo:** conforme al estudio estos tienen un impacto negativo en la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia, en la libertad de opinión e información, en el derecho a la libertad de asociación y en el derecho al trabajo deseado y de sindicato.

- **Sistemas de Inteligencia Artificial en el sistema de salud para realizar diagnósticos médicos:** tienen un impacto negativo en la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia.
- **Sistemas de Inteligencia Artificial en el sistema educativo para la calificación de ensayos:** tienen un impacto negativo en la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia.
- **Sistemas de Inteligencia Artificial para moderar contenido online e implementar los estándares de participación propuestos:** tienen un impacto negativo en la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia y en la libertad de opinión e información.
- **Sistemas de Inteligencia Artificial en los departamentos de recursos humanos para el reclutamiento y selección de candidatos:** conforme al estudio estos tienen un impacto negativo en la libertad de interferencias con la privacidad, la familia, la vivienda y la correspondencia, en la libertad de opinión e información y en el derecho a la libertad de asociación.



(Raso, Hilligoss, Krishnamurthy, Bavitz, & Kim, 2018)

Esto no significa que la tecnología *per se* tenga un impacto negativo en los derechos humanos descritos. Muchos de los resultados obtenidos están relacionados con la forma como los sistemas fueron implementados en cada uno de estos sectores y la manera como se interpretó la información proporcionada por esta tecnología. Este estudio provee entonces evidencia relevante sobre los potenciales riesgos que existen, pero no deben llevar a una generalización ni a considerar juicios absolutos, tales como que el uso de esta innovación de esta tecnología siempre tendrá efectos negativos en el sector de la salud o el sistema criminal. Es claro que pueden existir sectores y prácticas en los que la tecnología puede conllevar mayores riesgos, pero este debe ser evaluado caso a caso y bajo contextos particulares. El estado de la evidencia actual no permite generalizaciones al respecto.

Así mismo, existe evidencia en América Latina sobre la forma como la inteligencia artificial puede tener efectos no deseados. En este caso vale la pena también destacar el trabajo de la Web Foundation (World Wide Web Foundation, 2018) que analizó el uso de sistemas de inteligencia artificial por parte de las autoridades de Argentina y Uruguay.

El gobierno de la Provincia de Salta, Argentina, implementó un sistema para predecir embarazos adolescentes y deserción escolar, con el apoyo de Microsoft. El caso ilustra cómo un gobierno con recursos limitados busca usar tecnología para solucionar problemas sociales urgentes. El gobierno implementó un mecanismo para coordinar la recolección de datos de 200.000 personas que viven en poblaciones vulnerables a través de ONG y oficiales del gobierno, junto con un modelo de *machine learning* para generar predicciones sobre la deserción escolar y el embarazo adolescente dentro de miembros de esta población

La implementación desencadenó un gran interés por parte de otros gobiernos y críticas de activistas que consideraron que esto violaba la privacidad y no solucionaba las causas del problema. La implementación contaba con fases transparentes y otras menos transparentes u opaca. Mientras el gobierno no recolecte y consolide información sobre el impacto de estas herramientas, no es posible determinar si su implementación tiende a arrojar resultados justos o injustos (World Wide Web Foundation, 2018). Activistas de derechos de las mujeres han cuestionado la decisión de implementar este tipo de herramientas sin un marco que las incorpore en una política dirigida a la inequidad estructural que sufren las poblaciones a las que dice apoyar.

En Uruguay por su parte, el gobierno adquirió Predpol, un software policivo para predecir crímenes (World Wide Web Foundation, 2018). Es un caso problemático debido a su bajo grado de transparencia y a las dinámicas de discriminación y exclusión que puede reforzar. En menos de tres años el Ministerio del Interior descontinuó el programa y lo reemplazó con herramientas estadísticas retrospectivas, desarrolladas por el propio equipo del ministerio, que se consideraron más útiles.

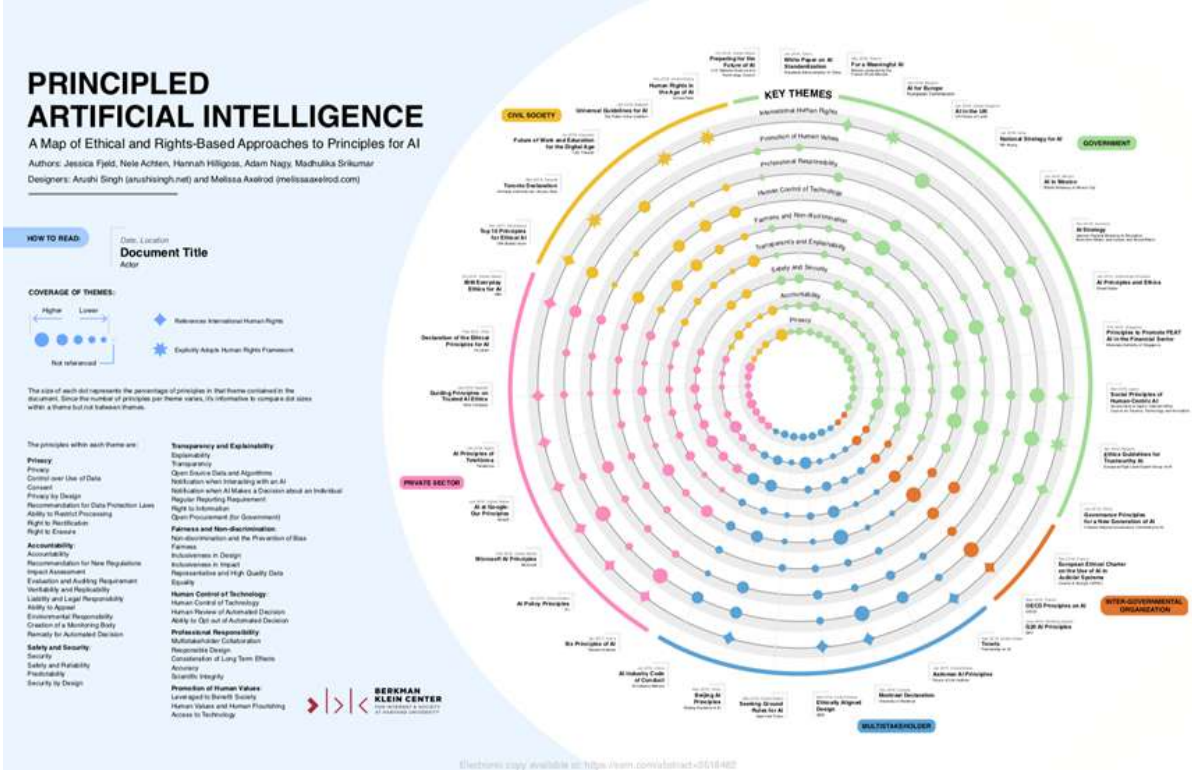
En este caso hay un riesgo de discriminación, y organizaciones locales e internacionales han alegado que herramientas como PredPol tienden a replicar los sesgos de entrenar datos y las dinámicas históricas de poder entre las fuerzas de orden público y poblaciones minoritarias o menos favorecidas y que se usan para justificar la presencia de la policía en áreas marginales (World Wide Web Foundation, 2018).

Colombia claramente debe evitar los problemas que ya se han descrito incluso en otros países de la región y los cuestionamientos éticos que se han generado al respecto. Todo lo anterior demuestra la necesidad tener de unos principios éticos que guíen el diseño, implementación y uso de esta tecnología. Estos deberían ser principios que aborden varios de los problemas específicos ya descritos y que tengan la posibilidad de ir cambiando y evolucionando conforme a los cambios que experimenta la tecnología. La implementación de estos principios también resulta ser una tarea prioritaria para ir mitigando varios de los riesgos que tiene el uso escalonado de esta innovación en distintos sectores de la sociedad, mientras aprendemos más de su funcionamiento y obtenemos mayor evidencia que permita considerar y justificar una mayor intervención del Estado al respecto.

IV
LOS PRINCIPIOS ÉTICOS PROPUESTOS PARA LA IA

Como ya se ha señalado, los principios éticos de la inteligencia artificial se han venido desarrollando en los últimos años y por ende resulta fundamental que el gobierno colombiano considere los principios ya existentes y la forma como pueden impactar al país y su ecosistema de innovación. En la actualidad incluso hacer un mapeo de estos principios ya no resulta necesario ya que hay disponibles insumos al respecto, como la investigación realizada por el Berkman Klein Center y publicada en enero de 2020, que presenta un mapeo completo de los principios de la Inteligencia artificial que ya se han propuesto a nivel mundial por parte de diversas entidades como gobiernos, organizaciones internacionales, ONGs y actores privados, entre otros. A continuación, se presentan dichos principios y las implicaciones de la forma cómo se ha abordado a nivel mundial:

Varios elementos de análisis se desprenden de este estudio:



(Fjeld, Achten, Hilligoss, Nagy, & Srikumar, 2020)

1 Los principios predominantes a nivel internacional

Como se puede observar, los principios que han predominado en el escenario mundial son los de privacidad, responsabilidad, seguridad, transparencia y explicación, justicia y no discriminación, control y supervisión humana de la

tecnología, promoción de los valores humanos y los señalados dentro marco internacional de derechos humanos.

La privacidad se constituye como un principio predominante y que llama la atención de todos los sectores analizados en este estudio. Esto se explica dada la gran cantidad de datos e información que hoy en día se utiliza para el diseño, desarrollo e implementación de esta tecnología. Los datos no son solo fundamentales para el funcionamiento de estos sistemas, sino que también hacen parte de prácticas conocidas como el entrenamiento de la inteligencia artificial. Las bases de datos para entrenamiento son comunes en la actualidad para el desarrollo de estos sistemas, especialmente de sistemas de *deep learning*. La forma como estos datos y su procesamiento pueden afectar la privacidad de los individuos constituye una preocupación recurrente a nivel internacional, punto que será profundizado más adelante.

Así mismo, la seguridad se ha constituido en una preocupación transversal considerando los distintos ataques que se han producido a esta tecnología, tales como el *adversarial machine learning*, y el impacto que un ataque de estas características puede llegar a tener. Esto no solo preocupa a los gobiernos, sino a los actores privados y grupos sociales que representan a los usuarios y consumidores.

Uno de los principios predominantes que genera mayor discusión es el que ha sido catalogado como el de control humano, dadas las implicaciones que puede tener este principio en el desarrollo de esta tecnología y las discusiones que puede generar. Esto es de especial preocupación para los gobiernos y la sociedad civil que temen el escalamiento y despliegue de sistemas de inteligencia artificial que no tenga ningún tipo de control y que tengan mayor autonomía en sus decisiones. Al contrario, el sector privado puede verse más inclinado a facilitar la autonomía de estos sistemas y tener límites claros a la intervención y control por parte de los humanos. Como se verá más adelante, este principio tiene distintas implicaciones y no sólo significa que los seres humanos controlan las decisiones y resultados que los sistemas reflejan, sino que esta tecnología debe privilegiar los valores humanos y su diseño debe reflejar las características propias de la humanidad, lo cual puede ser algo complejo y que debe ser visto con precaución.

Un principio muy ligado a la ética de las prácticas es el de responsabilidad profesional, el cual es de especial importancia y prevalencia en el sector privado y por las entidades que agrupan a distintos actores (*multistakeholder*). Para estas entidades la responsabilidad de los individuos detrás del desarrollo de esta tecnología es esencial y se busca que haya consensos al respecto.

2 Muchas de las preocupaciones están alrededor del uso de los datos

La protección de la privacidad puede considerarse como la mayor preocupación sobre la cual existe casi un consenso entre los distintos sectores de la sociedad. La capacidad de perfilamiento a partir del uso de datos o la toma de decisiones basados en información personal de baja calidad hace que la privacidad sea una preocupación horizontal.

Cabe preguntarse cuál es la relación entre este principio y las regulaciones de protección de datos e incluso muchas otras que protegen directamente la privacidad. Si ya existen regulaciones al respecto, no sería comprensible que este sea visto solo como un principio ético. Ahora bien, la privacidad como principio busca que se proteja la intimidad de las personas, pero entiende que esta protección puede ir más allá de los datos y de la forma como tradicionalmente se ha conceptualizado y regulado su protección. Considerar que los marcos legales existentes deben adaptarse a esta tecnología desconoce que ésta tiene un efecto disruptivo que hace que incluso estas normas deban ser repensadas. Por lo tanto, la privacidad como principio busca que dicha protección de la intimidad se siga privilegiando, sin importar el modelo legal que al final resulte.

Asimismo, este principio se pregunta por la forma como la inteligencia artificial y su implementación puede afectar la privacidad de un sujeto, no sólo desde el punto de vista del uso de sus datos, sino desde sus libertades fundamentales, su capacidad de decisión y de libre autodeterminación.

Ahora bien, este principio es predominante dadas las características actuales de esta tecnología. Sin embargo, puede que el desarrollo de nuevos algoritmos signifique un uso reducido de datos que transforme por completo nuestra comprensión de esta tecnología. Por lo tanto, la conceptualización de este principio ético no debe estar intrínsecamente ligado a los datos, y en especial a los datos personales.

3 Los gobiernos han privilegiado la transparencia

La transparencia se ha constituido como un principio predominante dentro de las propuestas de los gobiernos, algo que obedece a un fenómeno más extenso. En

los últimos años, el buen gobierno se ha visto relacionado con prácticas transparentes y con transparencia en el acceso a la información. La transparencia de la inteligencia artificial se une a estos esfuerzos por parte de distintos gobiernos y permite tener acceso a más información sobre la forma como las entidades públicas están tomando decisiones y la forma como están utilizando la tecnología. Por lo tanto, no es extraño ver que este sea un principio predominante en los gobiernos y en las distintas estrategias y documentos oficiales que señala la Universidad de Harvard en su reporte. Es claro que el uso de inteligencia artificial para el diseño de política públicas o en la prestación de servicios públicos estará mediado por dicho principio.

4 El sector privado busca definir responsabilidades en el uso de estos sistemas

Por su parte, parece que el sector privado tiene un especial interés en determinar las responsabilidades (*accountability*) que se derivan del diseño e implementación de esta tecnología. Dadas las implicaciones legales económicas y sociales que esto puede tener, los actores privados tienen un interés por establecer un entendimiento claro de este principio y sus limitaciones. Distribuir las responsabilidades que pueden derivarse del despliegue de esta tecnología puede tener profundos efectos económicos en el diseño de su producto, en el desarrollo del modelo de negocio e incluso en los modelos de contratación a utilizar con sus usuarios. Por lo tanto, es claro que los actores privados quieren influir en la forma como se entenderá este principio y sus implicaciones.

5 A la sociedad civil le preocupa la protección de derechos fundamentales y la discriminación

El mapeo realizado por la Universidad de Harvard demuestra que varios grupos representativos de la sociedad civil están interesados en la aplicación del marco de derechos humanos y la forma como éste se verá plasmado en la ética de la Inteligencia artificial. Esto no significa que sea el único sector que tiene esta preocupación, pero sí uno de los que más relevancia le ha dado a dicho marco. Específicamente estas preocupaciones giran alrededor de la justicia y la discriminación que se puede presentar a través de esta tecnología.

Unido a lo anterior, varios grupos de la sociedad civil ahondan en la necesidad de que exista un mayor control humano de esta tecnología y les preocupa la forma como la tecnología puede desplazar al ser humano en la toma de decisiones y en distintas actividades productivas. Las inquietudes frente a la discriminación se hacen presentes con los casos de estudio que se han realizado y que demuestran como distintos grupos social e históricamente marginados se han visto afectados por sistemas de inteligencia artificial. Los reportes elaborados por entidades que representan a la sociedad civil han hecho especial énfasis en este tipo de evidencia.

A esto se une la preocupación frente al uso que los gobiernos puedan valerse tecnología para generar políticas públicas y prestar servicios públicos y que pueden llegar a ser discriminatorias frente a un grupo específico, incluso sin tener este propósito. Ejemplo de esto es el uso de la inteligencia artificial en los sistemas de vigilancia y de lucha en contra del crimen que pueden tener un fin principal loable, pero que pueden resultar discriminatorios para cierto grupo social. El principio de no discriminación busca limitar estas consecuencias en el despliegue de esta tecnología.

El análisis del mapeo mundial de principios aplicables a la inteligencia artificial realizado por el Berkman Klein Center de la Universidad de Harvard permite **identificar cuáles son los intereses y preocupaciones que los principales sectores de la sociedad tienen frente a la selección de dichos principios y sus implicaciones.** Los reportes, documentos oficiales e investigaciones que los distintos actores han generado frente al tema permiten afirmar esto. No resultaría extraño que estas preocupaciones e intereses también se materialicen en las opiniones de los distintos actores que participen de las discusiones sobre el marco ético de la inteligencia artificial aplicable en Colombia. Por lo tanto, los principios que sean propuestos por el Gobierno Colombiano para dinamizar esta discusión deben considerar este contexto y las expectativas que distintos sectores sociales tienen al respecto.

V

ADOPCIÓN DE PRINCIPIOS ÉTICOS PARA LA IA EN COLOMBIA

A partir del mapeo de principios que se ha descrito y de la forma como distintos sectores de la sociedad a nivel mundial han abordado el tema, se sugieren los siguientes principios como aquellos que deben guiar el diseño, desarrollo, implementación y evaluación de los sistemas de inteligencia artificial en Colombia. Los principios se exponen de forma corta y sencilla, considerando que tienen la vocación de ser transversales y aplicables a la sociedad colombiana en general. Ahora bien, el efecto y el sentido que tiene cada uno de esto principios deben ser entendidos bajo el tipo de ética que se esté definiendo. Un principio como el de privacidad tendrá implicaciones distintas si se analiza desde la ética de los datos o la ética de los algoritmos. Así mismo, estos principios son descritos a través de un postulado general. Por lo tanto, el significado de cada uno de los principios se analiza bajo las tres éticas definidas anteriormente.

1 Transparencia

La transparencia debe ser entendida como la apertura para brindar información completa sobre el diseño, funcionamiento e impacto que tienen los sistemas de inteligencia artificial. Dicha información no debe comprometer la confidencialidad del modelo de negocio y de la innovación que la hagan susceptible de usos no autorizados.

En la ética de los datos: en la ética de los datos la transparencia tiene como efecto la necesidad de brindar información sobre las fuentes de la información que se utilizan para el diseño de esta tecnología, sus características y los fines para los cuales serán utilizados los datos, especialmente los datos personales. Esto tiene efectos en los datos utilizados para el entrenamiento e implementación de estos sistemas, dado que se debe ser transparente en su recolección, los criterios utilizados para su clasificación y procesamiento.

En la ética de los algoritmos: significa transparencia en el modelo detrás de los algoritmos, informado los criterios que llevan a estos sistemas a tener cierto tipo de resultados. En este punto es relevante brindar información a la ciudadanía sobre los insumos que se utilizan en el diseño de los sistemas y los resultados que puedan presentarse (*input and output transparency*). Así mismo, debe brindarse información completa sobre los falsos positivos o negativos que ha elaborado un algoritmo y los porcentajes de precisión (*algorithm accuracy*).

En ética de las prácticas: consiste en brindar información completa sobre los equipos y personas involucradas en el diseño y desarrollo de estos sistemas, los objetivos que persiguen y los manuales de conducta que hayan generado para desarrollar su trabajo. Estos manuales deben ser de acceso público. Como práctica para los encargados de la implementación de esta tecnología se debe privilegiar el uso de sistemas de código abierto, especialmente dentro de las entidades del Estado.

2 Explicación

Este principio es visto en momentos como complementario a la transparencia, pero se diferencia en que en la información que es compartida y dada a conocer a la ciudadanía de forma transparente debe ser comprensible tanto para los desarrolladores y usuarios del sistema como para aquellos individuos que pueden verse afectados por sus decisiones y resultados. Por lo tanto, va más allá de la transparencia y busca que el contenido de la información y la forma como se presenta sea de fácil acceso, considerando las complejidades que rodean a esta tecnología.

En la ética de los datos: la explicación hace que las personas puedan comprender la importancia de los datos en el diseño y desarrollo de estos sistemas, la forma como recolectan y procesan la información y los fines de hacerlo, en especial, cuando se procesan datos personales.

En la ética de los algoritmos: significa que los algoritmos puedan ser comprensibles en el sentido de que se permita entender los objetivos iniciales que busca y que son propios del modelo, y los resultados esperados y también los obtenidos. Es claro que los sistemas conocidos como de "*Black Box*" este principio puede entrar en tensión, dado que la capacidad de explicación puede llegar a ser limitada. Incluso los desarrolladores de este tipo de tecnología y expertos en la materia no logran entender en su totalidad el procesamiento que tiene lugar y la forma como los sistemas llegan a ciertos resultados. Sin embargo, lo que se busca es también explicar de forma comprensible cómo funcionan estos sistemas, porque son denominados como *Black Box* y las implicaciones que esto puede tener. Así mismo y, en cualquier caso, se debe presentar información comprensible y clara sobre los objetivos que persigue este sistema en su desarrollo e implementación.

En la ética de las prácticas: significa brindar información clara y precisa sobre los roles que tienen las personas involucradas en diseño, desarrollo e implementación de esta tecnología. Unido a lo anterior, debe generarse durante todo el proceso información clara, precisa y comprensible sobre la forma como estos sistemas están siendo evaluados y generar mecanismos específicos para compartir información sobre los resultados obtenidos, especialmente con las comunidades que se están viendo impactadas por estos sistemas.

3 Privacidad

La inteligencia artificial debe estar precedida de un respeto por la intimidad de las personas y su esfera privada que impide el uso información que estos no hayan autorizado y el perfilamiento de individuos a través de esta tecnología.

En la ética de los datos: en este campo el principio lleva a la necesidad de tener autorización para el uso de la información personal, cuando estos no sean datos públicos o bajo las excepciones que la ley específicamente señale, describiendo los fines y objetivos específicos perseguidos con el tratamiento (entrenamiento, funcionamiento, etc.). Igualmente, obliga a desarrollar mecanismos para mejorar la calidad de los datos utilizados y para lograr una actualización constante de la información. Esto lleva a que la población impactada tenga la posibilidad de corregir información personal equivocada o con errores que esté siendo utilizada para el desarrollo u operación de estos sistemas, sin que la funcionalidad de la tecnología pueda limitar este tipo de solicitudes.

En los algoritmos: el diseño de los algoritmos debe ser respetuoso de la intimidad de las personas y por ende los criterios de decisión no deben basarse en características personales y propias de su esfera privada. Debe limitarse el uso de información personal y solo usar aquella necesaria para un adecuado funcionamiento del sistema y que permita evitar los falsos positivos o negativos. Los diseñadores deben evitar el desarrollo de tecnologías que faciliten el perfilamiento de las personas, bajo criterios que no sean previamente conocidos y autorizados por estos. El uso de información para la mejora del funcionamiento o desempeño de estos sistemas debe ser informado a las personas que sean titulares de esa información.

En la ética de las prácticas: deben existir procedimientos internos que desarrollen buenas prácticas en el uso de información y en las respuestas y explicaciones que se brindan a los usuarios impactados por estas tecnologías. Debe privilegiarse las medidas de responsabilidad demostrada que permitan la implementación de herramientas de gestión de riesgo para la privacidad, al igual que los mecanismos de análisis de impacto en privacidad (*data protection impact assestments*). Los equipos de diseñadores y desarrolladores deben generar criterios para identificar aquellos casos en los que pueden presentarse perfilamientos, su impacto y la forma como pueden evitarse resultados negativos a partir de este proceso. Así no solo se protege la privacidad en un sentido individual, sino también colectivo, evitando generar clasificaciones o perfilamientos sociales no deseados.

4 Control Humano de las decisiones propias de un sistema de Inteligencia Artificial (*Human-in-the-loop* y *Human-over-the-loop*)

Este principio es aplicable a sistema de inteligencia artificial que tenga cierta autonomía en la toma de decisiones, haciendo que el ser humano tenga control total sobre la toma de decisiones, especialmente en una etapa de implementación (*Human-in-the-loop*). Una vez se haya alcanzado un mayor nivel de madurez de la tecnología en el país se pasará a que haya un mayor nivel de autonomía en la toma de decisiones, existiendo mecanismos de intervención de los seres humanos, especialmente cuando se presenten resultados no deseados. Dicha transición tendrá en cuenta el impacto social que pueda existir, en especial en el futuro del trabajo, dado el desplazamiento que se da de los seres humanos de ciertas actividades.

En la ética de los datos: la recolección y procesamiento debe ser realizada conforme a los parámetros y criterios establecidos por los seres humanos.

En la ética de los algoritmos: los algoritmos deben permitir y facilitar la toma de decisiones, pero en un principio deben servir de guía para la toma de decisiones y no pueden actuar de forma automatizada y conforme a modelos sugeridos.

En la ética de las prácticas: los sistemas de inteligencia artificial no deben ser utilizados para interactuar con la ciudadanía sin el control de un ser humano. Los sistemas de respuesta y conversación automatizado deben tener mecanismos para que los seres humanos intervengan y participen en cualquier momento. Debe evitarse prácticas que promuevan el relacionamiento con estos sistemas sin la verificación de que existen seres humanos detrás de los contenidos o respuestas que son generadas.

5 Seguridad

Los sistemas de inteligencia artificial no deben generar afectaciones a la integridad y salud física y mental de los seres humanos con los que interactúan.

En la ética de los datos: se deben implementar mecanismos que permitan asegurar que esta información mantendrá su confidencialidad, integridad y que en ningún momento puede verse alterada. Deben generarse mecanismos que puedan evitar este tipo de alteraciones a la información que estos sistemas utilizan y la forma como procesa la misma (*adversarial machine learning*).

En la ética de los algoritmos: la implementación de algoritmos y su diseño debe seguir un sistema de riesgos que permita establecer las posibles afectaciones que ciertos resultados pueden generar y la posibilidad de evitarlo. En ningún caso, un algoritmo debe llevar a un resultado que ponga en riesgo la integridad de un ser humano. Dichas decisiones solo pueden ser tomadas por seres humanos, y, en cualquier caso, los algoritmos servirán como modelos de guía en la toma de decisiones que los humanos tomen frente a la vida e integridad de otros (ejemplo: sector saludo o sector de seguridad nacional).

En la ética de las practicas: se deben evitar aquellas prácticas que pongan en riesgo a los sistemas de inteligencia artificial y los códigos de conducta deben generar parámetros para evitar aquellas actividades que en peligro la integridad y seguridad física de las personas.

6 Responsabilidad

Existe el deber de responder por los resultados que produzca un sistema de inteligencia artificial y las afectaciones que llegue a generar. Se partirá de la solidaridad en la responsabilidad de los diseñadores, desarrolladores y personas que implementen esta tecnología, por los daños que el uso de esta tecnología tenga en un individuo, salvo que se demuestre de forma suficiente que la responsabilidad recae en uno solo de estos actores.

En la ética de los datos: las entidades que recolectan y procesan datos para el diseño, desarrollo e implementación de sistemas de inteligencia artificial deben ser consideradas todas como responsables de esta información y deben responder por su integridad y las finalidades de procesamiento. En ningún caso la responsabilidad podrá recaer solo en uno de estos actores.

En la ética de los algoritmos: existe responsabilidad de las personas que diseñan un algoritmo por aquellos resultados que llegue a generar y los criterios utilizados para llegar a ciertas respuestas. Sin embargo, su responsabilidad no se deriva hasta la implementación, ya que en esta fase los resultados y su impacto serán responsabilidad de la persona o entidad que se encargue del uso de estos sistemas y de tomar decisiones a partir de los mismos.

En la ética de las prácticas: los involucrados en el desarrollo de esta tecnología deben establecer responsabilidades claras en la cadena de diseño, producción e implementación. Dentro de los equipos de trabajo debe existir una distribución clara de las funciones y de la responsabilidad en su desarrollo y cumplimiento. Se deben restringir prácticas y acuerdos entre actores que limiten la responsabilidad conforme a la forma como se ha establecido en este principio.

7 No discriminación

Los sistemas de inteligencia artificial no pueden tener resultados o respuestas que atenten contra el bienestar de un grupo específico o que limiten los derechos de poblaciones históricamente marginadas. Dichas decisiones solo podrán ser tomadas por seres humanos, bajo los criterios que el marco de derechos humanos permita en cada caso. La funcionalidad de un sistema de inteligencia artificial no debe estar limitado a un grupo específico por razón de sexo, raza, religión, discapacidad, edad u orientación sexual.

En la ética de los datos: los datos utilizados deben ser analizados de tal manera que se mitigue al máximo la posibilidad de usar información que contenga prejuicios o sesgos (*biases*), ya sea en su contenido, clasificación o el uso que se le ha dado. Se deben privilegiar mecanismos que permitan hacer un análisis previo de un conjunto de datos y de los posibles problemas que pueda tener.

En la ética de los algoritmos: los algoritmos deben ser capaces de responder a las necesidades e intereses de distintos grupos poblacionales. El adecuado desempeño de un algoritmo no puede verse limitado a un grupo poblacional específico. Debe existir un seguimiento constante de los falsos positivos y negativos que arroje un sistema estableciendo la forma como los criterios de sexo, raza, religión, discapacidad, edad u orientación sexual puedan afectar estos resultados.

En la ética de las prácticas: en el diseño debe participar un grupo diverso de la población y se deben generar matrices de impacto que permitan establecer de forma temprana algún tipo de discriminación y corregir los mismos oportunamente. Debe existir un análisis constante de dicho impacto e incluso considerar mecanismos para retirar inmediatamente sistemas que tengo efectos discriminatorios.

8 Inclusión

Es la participación activa de poblaciones históricamente marginadas en el diseño, desarrollo e implementación y evaluación de los sistemas de inteligencia artificial que se utilicen en Colombia. El Estado debe utilizar sistemas de inteligencia artificial que hayan cumplido con criterios de inclusión y respondan a las necesidades propias y específicas de estos grupos.

En la ética de los datos: esto implica la utilización de datos que sean representativos y procedentes de distintos grupos sociales, ya sea para el diseño, entrenamiento o funcionamiento de estos sistemas. Para este fin, se debe aumentar la disponibilidad de conjuntos de datos de los grupos históricamente menos representados.

En la ética de los algoritmos: las variables que han sido incluidas dentro del algoritmo reconocen los efectos que puede tener en contextos particulares y la posibilidad de que se privilegia a un grupo específico, evitando tal tipo de diseño.

En la ética de las prácticas: los grupos encargados del diseño, desarrollo e implementación deben tener en cuenta distintos sectores de la sociedad y se deben establecer comités de evaluación para evitar prácticas discriminatorias contra grupos como la mujer, afrodescendientes, indígenas o miembros de la comunidad LGBTI+, entre otros. Colombia debe liderar esfuerzos por evitar el diseño de sistema de inteligencia artificial que hagan que las mujeres sean vistas como asistentes personalizadas y como seres al servicio de los consumidores. El desarrollo de sistemas de inteligencia artificial que no tengan un género es deseable.

.

9 Prevalencia de los derechos de niños, niñas y adolescentes

Los sistemas de inteligencia artificial deben reconocer, respetar y privilegiar los derechos de niños niñas y adolescentes. En ningún caso está justificada la implementación de un sistema inteligente inicial que vaya en detrimento de su interés superior. Se debe abogar por fortalecer programas y estrategias de educación que faciliten el entendimiento de esta población de esta tecnología y facilite la interacción que estos tengan con esta innovación.

En la ética de los datos: los datos de esta población no pueden ser utilizados, salvo en aquellas actividades que se relacionen con su interés superior.

En la ética de los algoritmos: el diseño y desarrollo de los algoritmos debe ser comprensible para los niños, niñas y adolescentes, en especial cuando estos tengan un impacto sobre su desarrollo y bienestar. Se debe evitar el diseño de todo algoritmo que repercuta en un perjuicio para los menores de edad y en especial en prácticas como la intimidación y la discriminación (*bullying*).

En la ética de las prácticas: los niños, niñas y adolescentes deben ser considerados en el desarrollo de estos sistemas cuando sean propios de sus actividades, estableciendo mecanismos específicos de participación que además les permita evaluar el impacto que estos sistemas tienen en esta población. Se deben generar programas de capacitación y educación que les permita a los niños, niñas y adolescentes conocer y entender las características de esta tecnología y sus implicaciones resaltando la formación ética.

10 Beneficio social

Los sistemas de Inteligencia artificial que se implementan en Colombia deben permitir o estar directamente relacionada a una actividad que genere un beneficio social claro y determinable. Dicho beneficio puede verse materializado en la reducción de costos, el aumento de la productividad, la facilitación en la prestación de servicios públicos, entre otros. Los sistemas de inteligencia artificial que persigan otro tipo de fines no deben ser implementados en el sector público y se debe desincentivar su uso en otros sectores.

En la ética de los datos: el acceso fácil a los datos y la infraestructura de datos públicos deben ser priorizados para el desarrollo de sistemas de inteligencia artificial que muestran un claro beneficio social, en el diseño de política públicas y la prestación de servicios públicos.

En la ética de los algoritmos: los modelos y los algoritmos utilizados deben tener como fin último un resultado ligado a un fin socialmente reconocido, por lo que debe demostrarse como los resultados esperados se relacionan con dicho fin social.

En la ética de las prácticas: las personas que trabajan en el diseño, desarrollo e implementación de esta tecnología en Colombia deben conocer las principales dificultades sociales que enfrenta el país y establecer la forma como esta innovación y la implementación deseada puede ayudar a. resolverlo. El Estado debe promover el uso de esta tecnología dentro de un proceso de transformación digital que busca reducir brechas y disminuir la inequidad existente. Por esta misma razón, se deben establecer programas que promuevan los retos públicos en uso de inteligencia artificial (*IAckaton*) y que tenga como objetivo solucionar un problema social específico.

VI

HERRAMIENTAS PARA LA IMPLEMENTACIÓN DE LOS PRINCIPIOS PROPUESTOS

Una de las principales dudas que puede surgir al momento de analizar estos principios es la de establecer mecanismos específicos para su implementación y para que se vean materializados en el país y en el ecosistema de innovación y tecnología que buscan impactar. A continuación, se presentará una serie de mecanismos y herramientas concretas que facilitan este impacto, varias de las cuales ya han sido exploradas en otros países. Posteriormente, se mostrará de forma específica la forma como cada una de estas herramientas se relacionan con los principios descritos en este marco ético.

Sin embargo, ninguna de estas medidas asegura un éxito total en la implementación de estos principios o que los mismos se vean materializados de forma integral. Esta es una tarea compleja y frente a la cual se está en un estado de experimentación a nivel mundial. Así mismo, puede que sea necesario considerar nuevas herramientas y estrategias que no hayan sido descritas en este documento. Esto es algo completamente válido y deseable, dado que las herramientas que se describirán no constituyen una lista exhaustiva y taxativa al respecto. Igualmente, puede que una entidad considere que no es necesario aplicar todas estas medidas, sino algunas de ellas, conforme al contexto y las necesidades propias de cada entidad. Lo más importante en la implementación de este tipo de medidas es que logren materializar y plasmar los objetivos que se buscan con los principios ya descritos.

1 Evaluación de algoritmos (*algorithm assestment*)

Esta es una herramienta que se viene explorando en los últimos años por parte de distintas autoridades mundiales. Uno de los principales reportes al respecto fue elaborado por la autoridad de datos de Nueva Zelanda. A partir del reporte y análisis de casos de implementación de esta tecnología en entidades públicas de este país, la autoridad busca asegurarse de que los ciudadanos de Nueva Zelanda estén informados sobre el uso de algoritmos gubernamentales y de los pesos y contrapesos que existen para manejar su uso (Stats New Zealand, 2018). El reporte elaborado por el gobierno de Nueva Zelanda parte de los Principios para el uso seguro y efectivo de datos y análisis desarrollado por el Comisionado de Privacidad y el Guardián Jefe de Datos del Gobierno, y hace recomendaciones para mejorar la transparencia y responsabilidad (*accountability*) en el uso de algoritmos por parte del gobierno (Stats New Zealand, 2018). Los resultados del reporte brindan una oportunidad para que las agencias revisen y refresquen el proceso que utilizan para manejar algoritmos y ayudará a dar forma al trabajo del Guardián Jefe de Datos del Gobierno y del Oficial Digital Jefe del Gobierno, para promover innovación y buenas prácticas a través del sistema de datos (Stats New Zealand, 2018).

Dentro de los resultados del reporte y las recomendaciones realizadas se encuentra que, aunque algunas de las agencias del gobierno describen el uso de algoritmos con un estándar de buenas prácticas, no hay consistencia a lo largo del gobierno (Stats New Zealand, 2018). Hay oportunidades significativas para que las agencias mejoren las descripciones sobre cómo los algoritmos informan o impactan la toma de decisiones, en particular en aquellos casos donde hay un grado de toma automática de decisiones o donde los algoritmos soportan decisiones que tienen un impacto significativo en individuos o grupos (Stats New Zealand, 2018).

Adicionalmente, mientras algunas agencias del gobierno tienen procesos formales para revisar los algoritmos durante su desarrollo y operación, la mayoría no los tiene (Stats New Zealand, 2018). No hay consistencia a lo largo del gobierno para incluir estos procesos dentro de la policía organizacional, en vez de depender en la responsabilidad individual (Stats New Zealand, 2018). Esto sugiere que hay un amplio margen para mejorar, tanto para apoyar a quienes toman decisiones como para asegurarse de la continua mejoría de los algoritmos (Stats New Zealand, 2018).

La mayoría de agencias del gobierno indicó que esperan desarrollar en el futuro algoritmos operacionales que dependan de inteligencia artificial (Stats New Zealand, 2018). Va a ser retador explicar de manera clara cómo estos tipos de algoritmos funcionan y soportan la toma de decisiones y cómo se llega a un resultado determinado (Stats New Zealand, 2018). A medida que la tecnología evoluciona, esta seguirá siendo un área donde las agencias del gobierno deben balancear la importancia de la supervisión humana con posibles eficiencias en la prestación de servicios.

El gobierno colombiano debe considerar la elaboración de este tipo de análisis y reportes, como el realizado por el gobierno de Nueva Zelanda y cuyos resultados mencionamos anteriormente. Esto no solo le permitiría tener un mapeo constante de aquellos proyectos transformadores dentro del sector público en los que está utilizando esta tecnología, sino la forma como se están implementando principios como los señalados en este documento dentro de la implementación y despliegue de esa tecnología.

2 Auditoría de algoritmos (*algorithm auditing*)

Esta es una propuesta que han liderado varias entidades de la sociedad civil.

El Gobierno francés ha sido uno de los principales promotores de su implementación desde el sector público (Kayser-Bril, 2019). Incluso dentro de su estrategia nacional, el gobierno francés considero la creación de una plataforma nacional para la auditoría de algoritmos, especialmente los utilizados por el gobierno (Kayser-Bril, 2019). Sin embargo, esta propuesta sigue en discusión dentro del parlamento de ese país.

Muchos comportamientos algorítmicos que podríamos considerar antisociales pueden ser detectados a través de auditorías adecuadas, por ejemplo, explícitamente explorando el comportamiento de servicios a los consumidores y midiendo cuantitativamente resultados como discriminación por género en un experimento controlado (Kearns & Roth, 2020). Sin embargo, a la fecha, estas auditorías han sido llevadas a cabo principalmente de forma ad-hoc y aisladamente, usualmente por académicos o periodistas, y a menudo violando los términos de servicio de las compañías que están siendo auditadas (Kearns & Roth, 2020). En consecuencia, es necesario buscar maneras más sistemáticas, continuas y legales para auditar algoritmos (Kearns & Roth, 2020). Regular algoritmos es diferente y más complicado que regular la toma de decisiones por parte de los humanos (Kearns & Roth, 2020). Aquella regulación se debe basar en lo que hemos llamado diseño ético de algoritmos, que ahora está siendo desarrollado por una comunidad de cientos de investigadores (Kearns & Roth, 2020). El diseño ético de algoritmos empieza con un entendimiento preciso de los tipos de comportamiento que queremos que los algoritmos eviten (para saber qué buscar en una auditoría) y luego continúa con el diseño e implementación de algoritmos que eviten esos comportamientos (Kearns & Roth, 2020).

3 ‘Limpieza’ de datos (*data cleaning*)

Este tipo de medida busca limitar los prejuicios y los errores en los datos que se utilizan en el desarrollo de esta tecnología. Para esto, se han generado una serie de pasos que permiten un proceso de depuración, corrección y actualización de esta información, dentro de los que vale resaltar los siguientes:

1. *Monitorear los errores*: llevar un registro y observar las tendencias sobre dónde viene la mayoría de los errores, pues esto hará más fácil identificar y arreglar los datos incorrectos o corruptos (Gimenez, 2018). Esto es especialmente importante si se integran otras soluciones con el software administrativo principal, para que los errores no obstruyan el trabajo de otros departamentos (Gimenez, 2018).

2. ***Estandarizar los procesos***: es importante estandarizar el punto de entrada y revisar su importancia (Gimenez, 2018). Al estandarizar el proceso de datos, se asegura un buen punto de entrada y se reduce el riesgo de duplicación (Gimenez, 2018).
3. ***Validar la precisión***: validar la precisión de los datos una vez se haya limpiado la base de datos existente (Gimenez, 2018). Se recomienda investigar e invertir en herramientas de datos que ayuden a limpiar los datos en tiempo real (Gimenez, 2018). Algunas herramientas incluso utilizan IA o *machine learning* para mejorar la precisión (Gimenez, 2018).
4. ***Buscar datos duplicados***: identificar los duplicados puede ayudar a ahorrar tiempo al analizar los datos (Gimenez, 2018). Esto puede evitarse buscando y utilizando las herramientas de limpieza de datos mencionadas anteriormente, que pueden analizar datos en masa y automatizar el proceso (Gimenez, 2018).
5. ***Analizar***: Una vez los datos han sido estandarizados, validados y revisados por duplicados, se deben usar terceros para agregar los datos (Gimenez, 2018). Las fuentes externas confiables pueden recolectar información de primera mano, luego limpiar y compilar los datos para proveer información más completa para inteligencia y analítica de negocios (Gimenez, 2018).
6. ***Comunicarse con el equipo***: Comunicar el nuevo proceso estandarizado de limpieza al equipo. Ahora que se han limpiado los datos, es importante mantenerlos así (Gimenez, 2018). Esto ayudará a desarrollar y fortalecer la segmentación de consumidores y enviar información mejor dirigida a los consumidores y prospectos, por lo cual es importante que todo el equipo esté en la misma página (Gimenez, 2018).

Esta es una medida especialmente relevante en aquellos sectores que pueden ser susceptible al uso de datos que pueden tener mayores prejuicios o que puede estar "contaminada". Así mismo, es una medida altamente recomendable cuando las entidades se encuentren ante bases de datos de cuya calidad existan varias dudas.

4 Explicación inteligente

Como ya se ha señalado, el principio de explicación es uno de los presenta mayores desafíos en lo relacionado con su materialización dada la complejidad de varios sistemas de inteligencia artificial y que en algunos casos su funcionamiento no es del todo comprensible, incluso para expertos en la materia. Por esto se ha propuesto un modelo que podemos considerar como de 'explicación inteligente' que busca que dicha explicación aporte información comprensible a la ciudadanía sobre esta innovación desde que exista un análisis de costo-beneficio que justifica esta

medida. En este caso se reconoce que la explicación puede ser dispendiosa y costosa, por lo que la misma sólo debería proceder en aquellos casos específicos en que el acceso a este tipo de información presente más beneficios que costos. Es por esto que se considera como una explicación inteligente.

Así, se debe pensar sobre por qué y cuándo las explicaciones son lo suficientemente útiles como para superar los costos (Doshi-Velez & Kortz, 2017). Requerir que todos los sistemas de AI expliquen todas las decisiones puede resultar en sistemas menos eficientes, decisiones de diseño forzadas y un sesgo hacia resultados explicables pero insuficientes (Doshi-Velez & Kortz, 2017). Por ejemplo, los sobrecargos de forzar a una tostadora a explicar por qué cree que el pan está listo puede prevenir a una compañía de implementar una característica de tostadora inteligente, debido a los retos de ingeniería o preocupaciones por repercusiones legales (Doshi-Velez & Kortz, 2017). Por otra parte, podemos estar dispuestos a aceptar los costos económicos de un sistema de aprobación de créditos más explicable pero menos preciso por el beneficio social de ser capaces de verificar que no sea discriminatorio (Doshi-Velez & Kortz, 2017). Hay unas normas de la sociedad sobre cuándo necesitamos explicaciones y estas normas se deben aplicar a los sistemas de IA también (Doshi-Velez & Kortz, 2017).

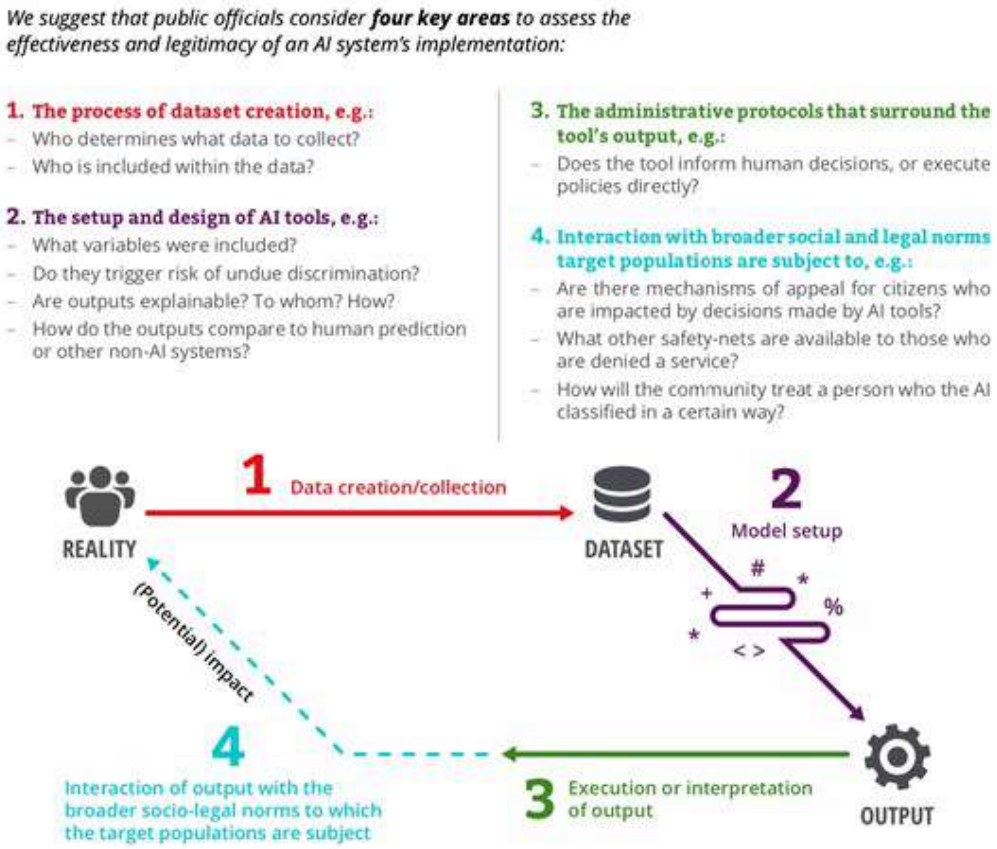
Al hacer esto, prevenimos que los sistemas de IA tengan pases libres para evitar que tengan el mismo nivel de escrutinio que pueden tener los humanos, y también evitamos pedir mucho de los sistemas de IA al punto de obstruir la innovación y el progreso (Doshi-Velez & Kortz, 2017). Incluso este paso modesto tendrá sus retos, y mientras los resolvemos obtendremos una mejor noción sobre si y dónde los requerimientos de explicación deben ser diferentes para los sistemas de IA y para los humanos (Doshi-Velez & Kortz, 2017). Dado que tenemos pocos datos para determinar los costos reales de requerir que los sistemas de IA den explicaciones, el rol de la explicación en asegurar la responsabilidad también debe reevaluarse de tiempo en tiempo, para adaptarse al panorama de la siempre cambiante tecnología (Doshi-Velez & Kortz, 2017).

5 Evaluación de la legitimidad

La Web Foundation ha desarrollado un modelo para **evaluar la legitimidad en la implementación de los sistemas de inteligencia artificial, especialmente por parte de las entidades públicas.** La legitimidad en la implementación se da cuando el procedimiento es explicable y tiene unas responsabilidades que se pueden rastrear, permitiendo definir con precisión quiénes son los involucrados en las

distintas operaciones del diseño y desarrollo de un sistema de inteligencia artificial, unido a unos resultados que no son discriminatorios, que son justos y en los que se puede determinar y minimizar el impacto de los falsos positivos y negativos (World Wide Web Foundation, 2018).

Ahora bien, para determinar si esa legitimidad se está presentando se proponen cuatro pasos específicos que deben seguir las entidades que están implementando esta tecnología de forma previa:



(World Wide Web Foundation, 2018)

6 Diseño de sistemas sostenibles y confiables

Para asegurarse de que un sistema de inteligencia artificial funciona de manera segura, se deben priorizar los objetivos técnicos de precisión, confiabilidad, seguridad y solidez (Leslie, 2019). Esto requiere que el equipo técnico examine previamente a profundidad cómo construir un sistema que opere precisa y confiablemente,

conforme a las expectativas de sus diseñadores, incluso cuando se confronta con cambios inesperados, anomalías y perturbaciones (Leslie, 2019). Construir un sistema de IA que cumpla estas metas de seguridad también requiere realizar rigurosamente pruebas, validaciones y reevaluaciones, así como la integración de mecanismos de supervisión y control adecuados en su operación en el mundo real (Leslie, 2019).

En *machine learning*, la **precisión** de un modelo es la proporción de ejemplos por los que genera un resultado correcto (Leslie, 2019). Esta medida de desempeño también se caracteriza a veces como una tasa de error o la fracción de casos por los que el modelo produce un resultado incorrecto (Leslie, 2019). Se debe tener en cuenta que en ocasiones la elección de una tasa de error o de un nivel de precisión aceptable se puede aceptar de acuerdo con el uso de las necesidades específicas de la aplicación (Leslie, 2019). En otros casos, se puede determinar por un estándar previamente establecido (Leslie, 2019).

Como una medida de desempeño, la precisión debe ser un componente central para establecer y matizar el acercamiento de un equipo a IA segura (Leslie, 2019). Especificar un nivel de desempeño razonable para un sistema a menudo requiere el perfeccionamiento o cambio de la medida de precisión (Leslie, 2019). Por ejemplo, si ciertos errores son más significativos o costosos que otros, una métrica para el costo total se puede incluir en el modelo para que el costo de algunos errores pueda ser superior que otros (Leslie, 2019). Así mismo, si la precisión y sensibilidad del sistema para detectar eventos raros es una prioridad, se puede usar la técnica de precisión y memoria (Leslie, 2019). Este método para abordar clasificaciones no balanceadas permite pesar la proporción de las detecciones correctas del sistema contra la proporción de detecciones reales del evento raro (Leslie, 2019).

En general, medir la precisión en la incertidumbre es un reto al que se le debe dar una atención significativa (Leslie, 2019). El nivel de confianza en el sistema de inteligencia artificial dependerá bastante en los problemas inherentes a intentos para modelar la realidad cambiante y caótica (Leslie, 2019). Las preocupaciones sobre la precisión deben afrontar temas de ruido inevitable que va a estar presente en los datos de muestra, las incertidumbres arquitectónicas generadas por la posibilidad de que a un modelo dado le falten características relevantes de la distribución subyacente y los cambios inevitables en los datos a medida que pasa el tiempo (Leslie, 2019).

Por otra parte, el objetivo de la **confiabilidad** es que un sistema de IA se comporte exactamente como sus diseñadores pretendieron y anticiparon (Leslie, 2019). Un sistema confiable se adhiere a las especificaciones para las que estaba programado llevar a cabo (Leslie, 2019). La confiabilidad es entonces una medida de consistencia

y puede determinar la confianza en la seguridad de un sistema, con base en la credibilidad con la que su operación se conforme con la funcionalidad pretendida (Leslie, 2019).

Por su parte, la meta de la **seguridad** abarca la protección de varias dimensiones operacionales de un sistema de IA cuando se confronta con un posible ataque adversarial (Leslie, 2019). Un sistema seguro es capaz de mantener la integridad de la información que lo constituye (Leslie, 2019). Esto incluye proteger su arquitectura de modificaciones no autorizadas o de daños a cualquiera de sus partes (Leslie, 2019). Un sistema seguro también debe ser continuamente funcional y accesible a sus usuarios autorizados, manteniendo la información privada y confidencial segura incluso bajo condiciones hostiles y adversariales (Leslie, 2019).

Finalmente, el objetivo de la **solidez** puede pensarse como el propósito de que un sistema de IA funcione de manera confiable y precisa bajo condiciones duras, que pueden incluir intervención adversarial, errores de quien lo implemente o ejecuciones distorsionadas por un aprendiz automatizado (Leslie, 2019). La medida de la solidez es en consecuencia la fuerza de la integridad de un sistema y la congruencia de su operación en respuesta a condiciones difíciles, ataques adversariales, perturbaciones, envenenamiento de datos y su comportamiento de aprendizaje reforzado indeseable (Leslie, 2019).

Dentro de las medidas a implementar se encuentran, según Leslie (2019):

1. Correr simulaciones extensivas durante la etapa de pruebas, para que las medidas apropiadas de restricción puedan programarse en el sistema.
2. Inspeccionar y monitorear continuamente el sistema, para que su comportamiento pueda ser mejor previsto y entendido.
3. Encontrar maneras de hacer que el sistema sea interpretado más fácilmente, para poder evaluar mejor sus decisiones.
4. Cablear mecanismos en el sistema que permitan a los humanos invalidar y apagar el sistema,

Esto debe involucrar protocolos rigurosos para probar, validar, verificar y monitorear la seguridad del sistema, así como autoevaluaciones del desempeño de la seguridad de los sistemas de IA por miembros relevantes del equipo en cada etapa del flujo de trabajo (Leslie, 2019). Estas autoevaluaciones deben valorar cómo las prácticas de diseño e implementación del equipo concuerdan con los objetivos de seguridad de precisión, confiabilidad, seguridad y solidez (Leslie, 2019). La autoevaluación debe

constar en un único documento en una forma que permita la revisión y la reevaluación (Leslie, 2019).

Dos medidas que deben ser implementadas de forma obligatoria y previa al despliegue de esta tecnología en el país son *F-1 scores* (Shmueli, 2019) y *confusion matrix* (Narkhede, 2018). Ambas son metodologías que han permitido definir la precisión de los resultados obtenidos por uno de estos sistemas, dando la posibilidad para realizar mejoras tempranas en estos sistemas.

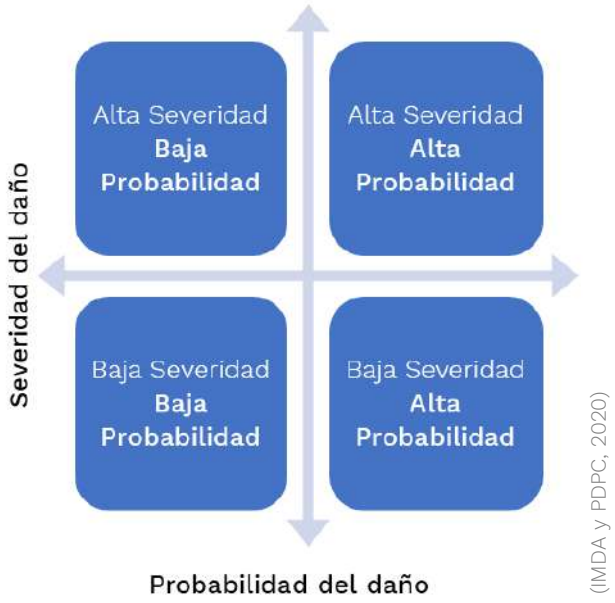
7 Definición y gestión de riesgos

Las organizaciones pueden considerar implementar un sistema coherente de gestión de riesgos y controles internos que afronten los riesgos involucrados en el despliegue de un determinado modelo de IA (IMDA y PDPC, 2020). Estas medidas incluyen:

1. Usar esfuerzos razonables para asegurarse de que las bases de datos utilizadas para entrenar el modelo de IA son adecuadas para el propósito pretendido, y evaluar y gestionar los riesgos de la imprecisión o los sesgos, así como revisar las excepciones identificadas durante el entrenamiento del modelo (IMDA y PDPC, 2020). Prácticamente, ninguna base de datos está totalmente sin sesgos (IMDA y PDPC, 2020). Las organizaciones deben propender por entender las formas en que las bases de datos pueden estar sesgadas y enfrentar esto en sus medidas de seguridad y estrategias de despliegue (IMDA y PDPC, 2020).
2. Establecer sistemas de monitoreo y reporte, así como procesos para asegurar que el nivel administrativo competente esté informado del desempeño y otros asuntos relacionados con el sistema desplegado (IMDA y PDPC, 2020). Cuando sea apropiado, el monitoreo puede incluir monitoreo autónomo para escalar efectivamente la supervisión humana (IMDA y PDPC, 2020). Los sistemas de IA pueden diseñarse para reportar el nivel de confianza de sus predicciones y características de explicabilidad pueden enfocarse en por qué el modelo obtuvo cierto nivel de confianza (IMDA y PDPC, 2020).
3. Asegurar una apropiada transferencia de conocimiento cuando haya cambios en personal clave relacionado con las actividades de IA, lo cual reducirá el riesgo de movimiento de personal creando una brecha en la gobernanza interna (IMDA y PDPC, 2020).
4. Revisar la estructura de gobernanza interna y las medidas cuando hay cambios significativos a la estructura organizacional o al personal clave involucrado (IMDA y PDPC, 2020).

5. Periódicamente revisar la estructura de gobernanza interna y las medidas para asegurar su relevancia y efectividad continuas (IMDA y PDPC, 2020).

El marco modelo propuesto por IMDA y PDPC también propone una matriz para clasificar la probabilidad y la severidad de daños a un individuo como un resultado de la decisión tomada por una organización sobre ese individuo (IMDA y PDPC, 2020). La definición de daño y la computación de probabilidad y severidad dependen del contexto y varían de sector en sector (IMDA y PDPC, 2020). Por ejemplo, el daño asociado con un diagnóstico incorrecto de una condición médica de un paciente va a depender del daño asociado con una recomendación de producto incorrecta (IMDA y PDPC, 2020).



(IMDA y PDPC, 2020)

Ahora bien, no solo los actores privados están considerando un enfoque de gestión del riesgo, sino también los gobiernos lo están haciendo. Por ejemplo, los miembros de la Unión Europea han señalado la ausencia actual de un marco europeo común (Comisión Europea, 2020). La Comisión Alemana de Ética de Datos ha pedido un sistema basado en el riesgo de cinco niveles de regulación, que iría de no tener regulación para los sistemas de IA más inocuos a una prohibición completa para los más peligrosos (Comisión Europea, 2020). Dinamarca recientemente desplegó el prototipo de un Sello de Ética de Datos (Comisión Europea, 2020). Malta introdujo un sistema de certificación voluntaria para IA (Comisión Europea, 2020). Si la Unión Europea falla en brindar una aproximación para todos los países que la integran, hay un riesgo real de fragmentación en el mercado interno, lo cual perjudicaría los objetivos de confianza, seguridad jurídica y penetración del mercado.

8 Privacidad diferencial (*differential privacy*)

El objetivo general de la privacidad diferencial es asegurar que diferentes tipos de análisis estadísticos no comprometan la privacidad y que la privacidad se preserve si, luego del análisis, el analizador no sabe nada sobre las características de una base de datos, para que la información que se haga pública no sea perjudicial para un individuo (Garg, 2019). Conforme a lo propuesto por Cynthia Dwork, la privacidad diferencial describe una promesa hecha por quien tiene los datos al titular de esos datos, según la cual el titular no se afectará al permitir el uso de sus datos en un estudio o un análisis, sin importar qué otros estudios, bases de datos o fuentes de información estén disponibles (Garg, 2019). Para definir la privacidad en el contexto de una base de datos simple, al remover una persona de la base de datos y la consulta no cambia, entonces la privacidad de esa persona estaría completamente protegida (Garg, 2019). Esto quiere decir que al remover la persona de la consulta se comprueba que no había una filtración de datos en el resultado de la consulta (Garg, 2019).

9 Códigos internos de conducta y/o ética

El objetivo de los códigos de conducta y/o ética es establecer los comportamientos esperados de quienes desarrollan, despliegan y utilizan tecnologías basadas en datos, para asegurar que todas las personas en esta cadena cumplan con los principios éticos para las iniciativas de datos: respeto a las personas, respeto a los derechos humanos, participación y responsabilidad por las decisiones (UK Department of Health and Social Care, 2019).

Los elementos mínimos que debe contener cada código son: (i) los principios que sigue cada institución o entidad; (ii) el alcance; (iii) si hay o no exclusividad, teniendo en cuenta que la exclusividad puede limitar los beneficios en diferentes sistemas; (iv) valor, teniendo en cuenta que los proveedores de tecnología pueden generar valor significativo de cualquier producto fuera del alcance del producto inicial, lo cual debe ser reconocido; (v) propiedad intelectual; (vi) responsabilidad; (vii) auditoría; (viii) sesgos y cómo prevenirlos, teniendo en cuenta que la amenaza más grande a la tecnología basada en datos es la presencia real o posible de sesgos, por lo cual cualquier acuerdo comercial debe identificarse, así como la forma en que se maneja, por quién y a expensas de quién, y (x) los diferentes roles (UK Department of Health and Social Care, 2019).

Adicionalmente, la Oficina del Gabinete del Reino Unido publicó un Marco Ético de Ciencia de Datos que busca ayudar a los investigadores a medida que los métodos de *big data* comienzan a ser usados en el sector público (ICO, 2017). Este Marco Ético incluye seis principios: (i) iniciar con necesidades de los usuarios y beneficio público claros; (ii) usar los datos y las herramientas que tengan la mínima intrusión necesaria; (iii) crear métodos de ciencia de datos robustos; (iv) estar alerta a las percepciones del público; (v) ser tan abierto y responsable como sea posible, y (vi) mantener los datos seguros (ICO, 2017).

Hay un rol en todo esto para los concejos o juntas de ética, tanto dentro de las entidades e instituciones como a nivel nacional (ICO, 2017). Una organización grande puede tener su propia junta ética, que puede asegurar que sus principios éticos se apliquen y puede hacer evaluaciones de situaciones difíciles como el equilibrio entre los intereses legítimos y los derechos de privacidad (ICO, 2017). Un elemento importante es la relación entre la junta ética y los empleados con responsabilidades sobre los datos y análisis, como el oficial de protección de datos (ICO, 2017).

10 Estrategias de educación e investigación en ética de la inteligencia artificial

El gobierno francés ha hecho de esta una de sus prioridades y por eso propone las siguientes medidas:

1. Fomentar la educación de todos los actores involucrados en cadenas de algoritmos (diseñadores, profesionales, ciudadanos) en materia de ética (CNIL, 2018).
2. Incrementar iniciativas para investigación en IA ética y lanzar una causa nacional en un proyecto de investigación de interés general (CNIL, 2018).

Por su parte, resulta interesante los proyectos que han desarrollado otras entidades con experiencia en la materia, como el MIT, que buscan que la ética de la inteligencia artificial entre de forma constructiva e innovadora a los currículos escolares. El objetivo último es habilitar a los estudiantes para que vean la inteligencia artificial como manipulable, desde un punto de vista técnico y societario, y empoderar a los estudiantes con herramientas para diseñar IA con la ética en mente (MIT Media Lab Staff, s.f.). El proyecto del MIT busca desarrollar un currículo de código abierto para estudiantes de bachillerato en el tema de inteligencia artificial (MIT Media Lab Staff, s.f.). Mediante una serie de lecciones y actividades, los estudiantes aprenden

conceptos técnicos y las implicaciones éticas que esos conceptos conllevan, como el sesgo algorítmico (MIT Media Lab Staff, s.f.). Durante el currículo, los estudiantes aprenden a pensar sobre los algoritmos como opiniones, les enseñan a considerar *stakeholders* directos e indirectos y se involucran en actividades de diseño para re-imaginar sistemas de inteligencia artificial familiares (MIT Media Lab Staff, s.f.).

11 Análisis de impacto en privacidad (*Privacy Impact Assessments*)

Un análisis de impacto en privacidad es una herramienta importante que puede ayudar a identificar y mitigar riesgos de privacidad antes de realizar el tratamiento de los datos personales (ICO, 2017). Bajo el Reglamento General de Protección de Datos de la Unión Europea (GDPR) es altamente probable que realizar un análisis de impacto en privacidad sea un requerimiento para la analítica de datos masivos que involucre el tratamiento de datos personales (ICO, 2017). Las características únicas de la analítica de datos masivos pueden hacer que algunos pasos de un análisis de impacto en privacidad sean más difíciles, pero estos retos se pueden superar (ICO, 2017).

12 Aproximación ética a los datos (*litmus test*)

Una aproximación ética al tratamiento de datos personales en el contexto del *big data* es una herramienta de cumplimiento importante (ICO, 2017). Las juntas éticas en el nivel organizacional y en el nacional pueden ayudar a evaluar elementos y asegurar la aplicación de principios éticos (ICO, 2017). Las aproximaciones éticas al uso de datos personales pueden ayudar a construir confianza con los individuos (ICO, 2017). Hay un rol para establecer los estándares de *big data* y promover las mejores prácticas a través de las industrias (ICO, 2017). En ocasiones, estos principios se condensan en un simple *litmus test* para recordar a los empleados que piensen en ellos cuando planeen nuevos usos de los datos, por ejemplo, si quisieran que los datos de un miembro de su familia se utilicen de cierta forma (ICO, 2017).

13 Portabilidad de los datos (*Personal data stores*)

El uso de *data stores* puede ayudar a solucionar asuntos de justicia y ausencia de transparencia al dar a los individuos un mayor control sobre sus datos personales (ICO, 2017). Los *data stores* pueden soportar el concepto de portabilidad de los datos, que hace parte del GDPR, en relación con el re-uso de los datos personales de un individuo bajo su control (ICO, 2017). Se ha sugerido que una forma de incrementar el control de un individuo sobre el uso de sus datos es a través de lo que usualmente se conoce como *personal data stores*, o a veces servicios de gestión de información personal (ICO, 2017). Estos son servicios prestados por terceros que contienen los datos de personas en su nombre y los ponen a disposición de organizaciones cuando así lo desean los individuos (ICO, 2017). Lo anterior además es una medida que puede ayudar a Colombia con su el cumplimiento de los principios generales de privacidad de la OCDE.

14 Fortalecer la ética en los negocios y los programas de empresa y derechos humanos

Si bien en general las compañías indican que operan de manera responsable o ética, diferentes índices, estándares y certificaciones brindan cierta seguridad para los *stakeholders* con referencia a si en efecto la compañía está haciendo lo que dice que está haciendo, es decir, si está cumpliendo con sus compromisos (Institute of Business Ethics, 2012). Estos mecanismos también pueden ser un incentivo para las empresas.

Respetar los derechos humanos es un criterio en varios índices y estándares, como el UN Global Compact, el estándar ISO 26000 para responsabilidad social y el Índice de Sostenibilidad de Dow Jones (Institute of Business Ethics, 2012). Los requerimientos del Índice FTSE4Good varían dependiendo de si las compañías operan en sectores y países con alto o bajo impacto en derechos humanos (Institute of Business Ethics, 2012). Para calificar para inclusión, las empresas de alto impacto deben implementar, entre otros: (i) una política de derechos humanos; (ii) entrenamiento en derechos humanos; (iii) responsabilidad y *accountability* en derechos humanos a nivel de la junta directiva; (iv) involucramiento de los *stakeholders* con las comunidades locales, y (v) mecanismos de monitoreo y reporte regular sobre las actividades y los progresos (Institute of Business Ethics, 2012).

Dado que este es un tema en constante desarrollo y en el cual actualmente se están diseñando guías de implementación para hacer que las empresas sean responsables por violaciones de derechos humanos, como pueden serlo los gobiernos, es esencial que el Gobierno colombiano esté evaluando permanentemente sus compromisos frente a los desarrollos de las organizaciones internacionales, como la ONU, sobre este tema. De esta manera, el Gobierno puede implementar diferentes mecanismos para que las empresas cumplan con los derechos humanos y mitiguen los riesgos y amenazas que las diferentes tecnologías implican para los derechos humanos.

15 Modelos de Gobernanza para asegurar la ética de la inteligencia artificial

A grandes rasgos, se ha propuesto un modelo de Marco de Gobernanza de IA que contiene guías sobre las medidas para promover un uso responsable de inteligencia artificial que las organizaciones deben adoptar en las siguientes áreas clave, según (IMDA y PDPC, 2020):

1. Adaptar estructuras de gobernanza internas y medidas para incorporar valores, riesgos y responsabilidades relacionadas con la toma algorítmica de decisiones (IMDA y PDPC, 2020).
2. Determinar el nivel de involucramiento humano en la toma de decisiones aumentada por IA (IMDA y PDPC, 2020).
3. Gestión de operaciones: considerando elementos en el desarrollo, selección y mantenimiento de modelos de IA (IMDA y PDPC, 2020).
4. Estrategias para comunicarse e interactuar con los *stakeholders* de una organización y el manejo de las relaciones con ellos (IMDA y PDPC, 2020).

Las organizaciones que adopten este modelo pueden considerar que no todos los elementos son relevantes, pues el modelo está diseñado para ser flexible y las organizaciones pueden adaptarlo para ajustarlo a sus necesidades y adoptando los elementos relevantes (IMDA y PDPC, 2020).

Dentro de los elementos con los que debe contar una entidad particular, se deben implementar los siguientes roles y responsabilidades claros para un despliegue ético de IA (IMDA y PDPC, 2020):

1. Responsabilidad por y vigilancia de las varias etapas y actividades involucradas en el despliegue de IA debe asignarse al personal y/o departamentos apropiados (IMDA y PDPC, 2020). De ser necesario y posible, se debe considerar establecer un órgano coordinador, que tenga experiencia relevante y representación apropiada de toda la organización (IMDA y PDPC, 2020).
2. El personal y/o los departamentos que tengan funciones internas de gobernanza de IA deben ser completamente conscientes de sus roles y responsabilidades, estar adecuadamente entrenados y contar con los recursos y la guía necesarios para que cumplan sus deberes (IMDA y PDPC, 2020).

Los roles y responsabilidades clave que pueden ser asignados incluyen:

1. Usar cualquier marco de gestión de riesgos y aplicar medidas de control de riesgo para: (i) evaluar y gestionar los riesgos de desplegar IA, incluyendo cualquier potencial impacto adverso para los individuos; (ii) decidir sobre el nivel apropiado de involucramiento en la toma de decisiones ayudada por IA, y (iii) manejar el modelo de entrenamiento de IA y el proceso de selección (IMDA y PDPC, 2020).
2. Hacer mantenimiento, monitoreo, documentación y revisión de los modelos de IA que han sido desplegados, con miras a tomar remedios en caso de ser necesarios (IMDA y PDPC, 2020).
3. Revisar canales de comunicación e interacciones con los *stakeholders* para brindar divulgación y canales de retroalimentación efectivos (IMDA y PDPC, 2020).
4. Asegurar que el personal relevante que lidia con sistemas de IA esté adecuadamente entrenado (IMDA y PDPC, 2020). Cuando aplique y sea necesario, el personal que trabaje e interactúe directamente con los modelos de IA puede necesitar estar entrenado para interpretar el output del modelo de IA y las decisiones, así como para detectar y gestionar sesgos en los datos (IMDA y PDPC, 2020). Otros miembros del personal cuyo trabajo requiera la interacción con el sistema de IA debe estar entrenado para al menos estar alerta de y sensible sobre los beneficios, riesgos y limitaciones al utilizar IA, para que sepan cuándo alertar a los expertos en la materia dentro de sus organizaciones (IMDA y PDPC, 2020).

VII

RELACIÓN ENTRE LOS PRINCIPIOS PROPUESTOS Y LAS HERRAMIENTAS DE IMPLEMENTACIÓN

A continuación, se presenta la forma como las herramientas de implementación anteriormente descritas interactúan con cada uno de los principios. Como se puede observar, algunas herramientas pueden facilitar la implementación de todos los principios, mientras que otras medidas tienen un impacto más dirigido a unos principios específicos:

	Transparencia	Explicación	Privacidad	Control Humano	Seguridad	Responsabilidad	No discriminación	Inclusión	Prevalencia de los derechos de niños, niñas y adolescentes	Beneficio social
Evaluación de algoritmos	X	X	X	X	X	X	X	X	X	X
Auditoría de algoritmos	X	X	X		X					
'Limpieza' de datos			X		X	X	X			
Explicación inteligente	X	X					X			X
Evaluación de la legitimidad	X			X			X	X	X	X
Diseño de sistemas sostenibles y confiables					X	X				
Definición y gestión de riesgos			X	X	X	X				
Privacidad diferencial			X		X					
Códigos internos de conducta y/o ética	X	X	X	X	X	X	X	X	X	X
Estrategias de educación e investigación en ética de la inteligencia artificial	X	X	X	X	X	X	X	X	X	X
Análisis de impacto en privacidad			X		X		X			
Läuens Test			X	X	X	X	X			
Portabilidad de los datos			X							
Fortalecer la ética en los negocios y los	X	X	X	X	X	X	X	X	X	X

RECOMENDACIONES

Como ya se ha señalado, este marco debe servir como un insumo para el debate nacional y una discusión más amplia frente al tema. Por lo tanto, señalamos una serie de recomendaciones que se deberían considerar para que se dé esta discusión y la posterior adopción de un marco ético para la inteligencia artificial tanto en el sector público como privado:

1. Priorizar la publicación de un marco ético para la inteligencia artificial dados los riesgos identificados y con anterioridad a que se presente un mayor despliegue de esta tecnología. Por lo tanto, el Gobierno debería resaltar la importancia de este marco ético y por qué debe ser priorizado, en especial en aquellas entidades que ya hayan implementado esta tecnología o estén próximos a hacerlo.

2. El Gobierno, en cabeza de la Consejería Presidencial para Asuntos Económicos y Transformación Digital y el Ministerio de Tecnologías de Información y Comunicaciones, debe auspiciar e impulsar un diálogo nacional con distintos sectores de la sociedad dado el efecto transversal que tiene un marco de estas características. Sin embargo, dicho diálogo debe darse sobre una base concreta que permita orientar la discusión y sus objetivos. Esta propuesta espera ser un insumo inicial en este propósito. Este diálogo debe involucrar actores nacionales y referentes internacionales.

3. Posterior a las discusiones que se lleven a cabo, deben realizarse ajustes al marco. Estos ajustes deben incorporar las principales conclusiones que arrojen estas discusiones.

4. Las entidades de Gobierno deben liderar una campaña de difusión del marco que permita su conocimiento por los distintos sectores de la economía y la sociedad en su conjunto para recibir los insumos de los distintos sectores.

5. Un marco ético de estas características debe partir de principios generales, capaces de adaptarse a los cambios tecnológicos que se vayan presentado. Los principios propuestos no deben dejar de lado las propuestas internacionales que ya se han realizado y sus implicaciones en distintos sectores.

6. Un marco ético de la inteligencia artificial no debe desconocer el impacto que los principios propuestos tendrá en cada uno de los momentos de la cadena algorítmica y en lo que ha sido denominado la ética de los datos, algoritmos y prácticas.

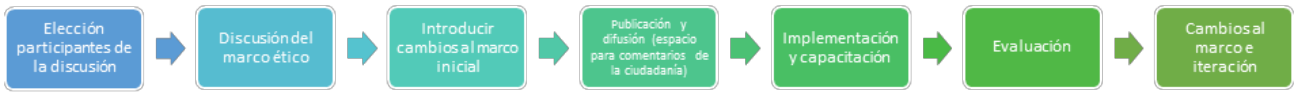
7 Las herramientas de implementación de los principios son esenciales y constituye uno de los elementos más relevantes del marco. Sin estos, el marco carece de herramientas para verse realmente materializados dentro de la sociedad colombiana. El gobierno debe ahondar en su diseño y forma de implementación, desarrollando un plan organizado para tal fin.

8 Unido a lo anterior, el proceso de implementación debe estar acompañado de una estrategia de sensibilización y capacitación en el sector público para que el marco sea entendido, en especial las herramientas de materialización que finalmente sean adoptadas.

9 Deben generarse métricas para evaluar los resultados y el impacto que tienen los principios propuestos. Este impacto debe medir el efecto que tiene el marco adoptado en la adopción de esta tecnología emergente y los casos de afectación a la ciudadanía y sus derechos fundamentales.

10 Aunque en un principio se espera que estos principios no deban ser adoptados mediante una normativa vinculante, el Gobierno debe analizar constantemente la necesidad de dar a conocer estos a través de algún instrumento legal y las características del mismo.

11 Dado lo novedoso del tema, es posible que surja una iteración constante de este proceso que permita un refinamiento de este marco, considerando los cambios y nuevos desafíos que puede traer una tecnología emergente en constante transformación como la inteligencia artificial.



Referencias:

AI and Inclusion Staff. (n.d.). *AI and Inclusion*. Retrieved from https://aiandinclusion.org/

Cabrol, M., González, N., Pombo, C., & Sanchez, R. (2020, Enero). *Adopción ética y responsable de la Intelidencia Artificial en América Latina y el Caribe*. Retrieved from Interamerican Development Bank: https://publications.iadb.org/publications/spanish/document/fAIr_LAC_Adopci%C3%B3n_%C3%A9tica_y_responsable_de_la_inteligencia_artificial_en_Am%C3%A9rica_Latina_y_el_Caribe_es.pdf

CNIL. (2018, Mayo 25). *Algorithms and artificial intelligence: CNIL's report on the ethical issues.* Retrieved from CNIL: https://www.cnil.fr/en/algorithms-and-artificial-intelligence-cnils-report-ethical-issues

Comisión Europea. (2020, Febrero 19). *White Paper On Artificial Intelligence - A European approach to excellence and trust.* Retrieved from European Commission: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., . . . Whitt. (2019, Diciembre). *AI Now Institute*. Retrieved from AI Now 2019 Report: https://ainowinstitute.org/AI_Now_2019_Report.pdf

Doshi-Velez, F., & Kortz, M. (2017). *Accountability of AI Under the Law: The Role of Explanation.* Retrieved from Berkman Klein Center Working Group on Explanation and the Law: http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372584

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020, Enero). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI.* Retrieved from Berkman Klein Center for Internet & Society: https://dash.harvard.edu/handle/1/42160420

Floridi, L., & Taddeo, M. (2016, Diciembre). What is Data Ethics? *Phil. Trans. R. Soc. A, 374*(2083).

Garg, A. (2019, Junio 11). *Differential Privacy and Deep Learning.* Retrieved from Medium: https://webcache.googleusercontent.com/search?q=cache:osZUXYGNnusJ:https://mc.ai/differential-privacy-and-deep-learning-2/+&cd=20&hl=en&ct=clnk&gl=co

Gimenez, L. (2018, Mayo 24). *6 steps for data cleaning and why it matters*. Retrieved from Geotab: https://www.geotab.com/blog/data-cleaning/

Giubilini, A. (2016, Marzo 14). *Conscience*. Retrieved from Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/entries/conscience/

ICO. (2017). *Big data, artificial intelligence, machine learning and data protection.* Retrieved from Information Commissioner's Office: https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

IMDA y PDPC. (2020). *Model Artificial Intelligence Governance Framework*. Retrieved from PDPC Singapore: https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf

Institute of Business Ethics. (2012, Julio). *Business Ethics Briefing Issue 26*. Retrieved from Business and Human Rights: https://www.business-humanrights.org/sites/default/files/media/business-ethics-human-rights-briefing_.pdf

Kayser-Bril, N. (2019, Enero 29). *Report Automating Society France.* Retrieved from Algorithm Watch: https://algorithmwatch.org/en/automating-society-france/

Kearns, M., & Roth, A. (2020, Enero 13). *Ethical algorithm design should guide technology regularion.* Retrieved from Brookings: https://www.brookings.edu/research/ethical-algorithm-design-should-guide-technology-regulation/

Leslie, D. (2019). *Understanding artificial intelligence ethics and safety A guide for the responsible design and implementation of AI systems in the public sector.* Retrieved from The Alan Turing Institute: https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf

MIT Media Lab Staff. (n.d.). *AI + Ethics Curriculum for Middle School*. Retrieved from MIT Media Lab: https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/

Narkhede, S. (2018, Mayo 9). *Understanding Confusion Matrix*. Retrieved from Towards Data Science: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Raso, F. A., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. (2018, Septiembre 25). *Artificial Intelligence & Human Rights: Opportunities & Risks.* Retrieved from Berkman Klein Center For Internet & Society at Harvard University: https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights

Shmueli, B. (2019, Julio 3). *Multi-Class Metrics Made Simple, Part II: the F1-score* . Retrieved from Towards Data Science: https://towardsdatascience.com/multi-class-metrics-made-simple-part-ii-the-f1-score-ebe8b2c2ca1

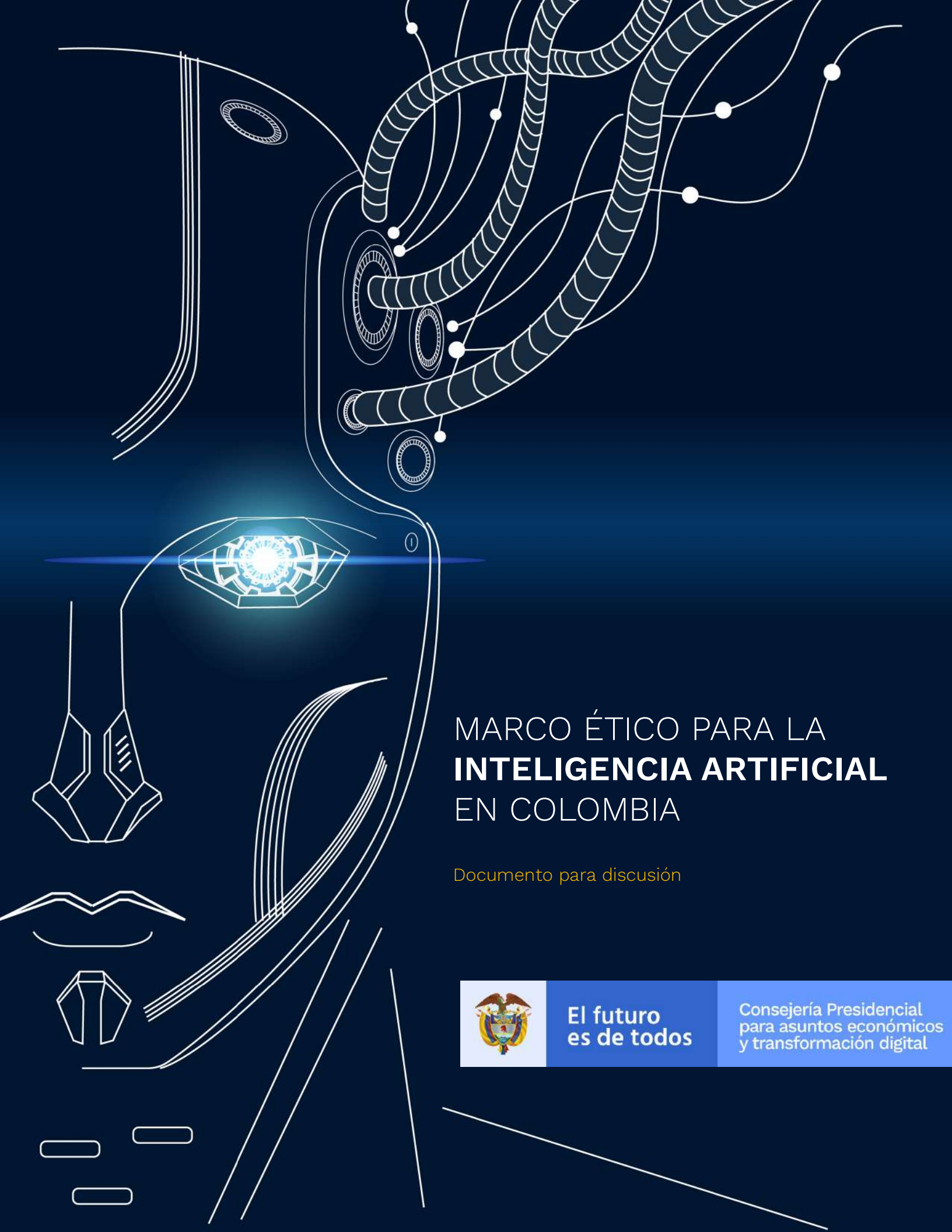
Stats New Zealand. (2018, Octubre). *Algorithm assessment report*. Retrieved from New Zealand Government: https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf

UK Department of Health and Social Care. (2019, Julio 18). *Code of conduct for data-driven health and care technology.* Retrieved from UK Department of Health and Social Care: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology

World Wide Web Foundation. (2017, Julio). *ALGORITHMIC ACCOUNTABILITY Applying the concept to different country contexts.* Retrieved from World Wide Web Foundation: https://webfoundation.org/docs/2017/07/Algorithms_Report_WF.pdf

World Wide Web Foundation. (2018). *HOW ARE GOVERNMENTS IN LATIN AMERICA USING ARTIFICIAL INTELLIGENCE? A proposal for effective and legitimate implementations of AI systems in the public sector.* Retrieved from World Wide Web Foundation: http://webfoundation.org/docs/2018/07/AI-in-Latin-America_Overview.pdf

World Wide Web Foundation. (2018, Septiembre). *World Wide Web Foundation.* Retrieved from ALGORITHMS AND ARTIFICIAL INTELLIGENCE IN LATIN AMERICA A Study of Implementation by Governments in Argentina and Uruguay: http://webfoundation.org/docs/2018/09/WF_AI-in-LA_Report_Screen_AW.pdf



MARCO ÉTICO PARA LA **INTELIGENCIA ARTIFICIAL** EN COLOMBIA

Documento para discusión

El futuro es de todos

Consejería Presidencial para asuntos económicos y transformación digital