



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



El futuro
es de todos

Gobierno
de Colombia



MARCO ÉTICO PARA LA INTELIGENCIA ARTIFICIAL EN COLOMBIA

Víctor Manuel Muñoz Rodríguez

Director Departamento Administrativo
de la Presidencia de la República

Cristhian Camilo Londoño

Consejero Presidencial para Asuntos
Económicos y Transformación Digital

Autores:

Armando Guío Español

Consultor CAF

Elena Tamayo Uribe

Contratista Presidencia de la República

Pablo Gómez Ayerbe

Asesor Consejería Presidencial de Asuntos
Económicos y Transformación Digital

Versión 1:

Mayo de 2021

Comentarios, sugerencias o correcciones
pueden ser enviadas al correo electrónico:
+ transformaciondigital@presidencia.gov.co



CONTENIDO





INTRODUCCIÓN

pág 7-18



OBJETIVO Y ALCANCE DEL DOCUMENTO

pág 19-21



ADOPCIÓN DE PRINCIPIOS ÉTICOS PARA LA IA EN COLOMBIA

pág 22-34



HERRAMIENTAS PARA LA IMPLEMENTACIÓN DE LOS PRINCIPIOS PROPUESTOS

pág 35-51



RECOMENDACIONES FINALES PARA EL SECTOR PÚBLICO

pág 52-60



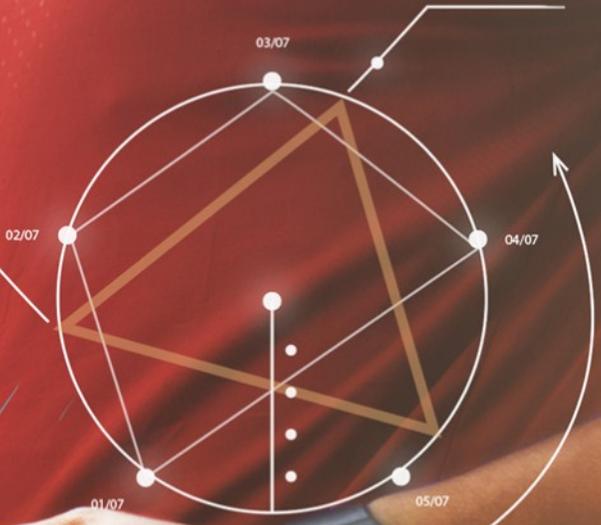


I INTRODUCCIÓN





Analysis
70 %



+

La cuarta revolución industrial está impactando de manera estructural la forma de vida de los ciudadanos, la toma de decisiones, los modelos económicos y las interacciones entre las empresas, el Estado y los ciudadanos. Esta revolución ha generado un uso extendido de la inteligencia artificial (IA), como tecnología de base para el desarrollo de diversos sistemas y soluciones con grandes impactos económicos y sociales. La inteligencia artificial ha permeado la mayoría de las industrias económicas como la manufactura, salud, agricultura y el arte, entre otros.

“La inteligencia artificial se entiende como un campo de la informática dedicado a resolver problemas cognitivos comúnmente asociados con la inteligencia humana o seres inteligentes, entendidos como aquellos que pueden adaptarse a situaciones cambiantes. Su base es el desarrollo de sistemas informáticos, la disponibilidad de datos y los algoritmos” (CONPES 3975).

La cuarta revolución industrial también está teniendo impactos considerables en Colombia y la transformación digital se ha visto acelerada en el contexto del COVID - 19. Como muestra lo ya enunciado, en lo corrido de 2021 se evidencia que los servicios de desarrollo de sistemas informáticos y procesamiento de datos han presentado un incremento del 25,8% en comparación al año 2019 (DANE, 2021). Asimismo, la apropiación digital pasó de 33% en febrero de 2020 a 55% en enero de 2021 (Centro Nacional de Consultoría, 2021). También se evidencian avances en la adopción de la inteligencia artificial y se han abierto nuevos ecosistemas como Health-tech, Fintech y Govtech, entre otros.

La cuarta revolución industrial y la inteligencia artificial en específico tienen el potencial de aumentar considerablemente el crecimiento económico y de resolver problemas sociales estructurales como el cambio climático, la desigualdad y la corrupción. El impulso de la inteligencia artificial podría aumentar el PIB de Colombia en un 6,8% en los próximos 10 años (Zaballos, 2021). Asimismo, se prevé que el mercado de software de inteligencia artificial (IA) en América Latina genere unos ingresos de 1.340

millones de dólares en 2021 (Navarro, 2021). Sin embargo la inteligencia artificial también puede generar efectos negativos en materia de desinformación, sesgos, discriminación, seguridad y afectaciones a la privacidad. Por consiguiente, surge una necesidad de asegurar una implementación ética y confiable de la inteligencia artificial en Colombia.

De la misma manera, basándose en los principios de inteligencia artificial de la OCDE, UNESCO, y en los principios destacados por los resultados de una investigación realizada por expertos de la universidad de Harvard, el Gobierno Nacional se percató que debía haber un impacto directo en sus entidades públicas y por ende ha priorizado el uso ético de la inteligencia artificial en estas entidades.

La Presidencia de la República ha identificado la importancia de las consideraciones éticas para el diseño, el desarrollo y la implementación de la inteligencia artificial en Colombia y más precisamente la necesidad de adoptar un marco ético como guía no obligatoria para la implementación de la inteligencia artificial en el país.

Con base en lo anterior, y con el objeto de facilitar la implementación de la Política Nacional para la Transformación Digital e Inteligencia Artificial (CONPES 3975), la Presidencia de la República de Colombia realizó el documento del Marco Ético para la Inteligencia Artificial en Colombia con base en el documento técnico elaborado por Armando Guio Español y aportado por la CAF (Banco de Desarrollo de América Latina). Asimismo, se han incorporado los insumos de las 6 mesas técnicas lideradas por diferentes entidades de orden nacional e internacional, entre las cuales se encuentran la CAF, el BID y la universidad de Harvard a través del Berkman Klein Center for Internet & Society.

La mesa redonda liderada por el Berkman Klein Center for Internet & Society se centró en los aspectos del documento técnico relacionados con la juventud. En concreto, se exploró cómo el Marco Ético de Colombia

- + puede garantizar que los jóvenes estén protegidos de los posibles impactos negativos de la IA, que puedan aprovechar al máximo las oportunidades de esta tecnología y tengan las habilidades necesarias para dar forma tanto al futuro desarrollo de las tecnologías de la IA como a las políticas en torno a las tecnologías.

En esta mesa redonda participaron más de 25 expertos en educación y juventud, políticas públicas de inteligencia artificial y elaboración de políticas públicas nacionales e internacionales. La mesa redonda comenzó con la retroalimentación sobre el Marco Ético y el principio "Prevalencia de los derechos de los niños y adolescentes (NNA)", la segunda parte estuvo orientada a las mejores prácticas para implementar el Marco Ético con base en la experiencia internacional. Los principales temas abordados en el transcurso de la conversación fueron: el dinamismo y la contextualización, el equilibrio entre la protección, la preparación y el empoderamiento, la perspectiva de múltiples partes interesadas y los enfoques legales para la implementación.

Principales recomendaciones de los participantes para tener en cuenta en la construcción del Marco Ético:

- + **Incorporar el concepto de dinamismo en el principio.** El principio debe ser dinámico para adaptarse a los avances tecnológicos, especialmente de cara a las nuevas generaciones.
- + **Incorporar el concepto de contextualización en el principio.** La contextualización debe reflejar las diferentes percepciones de los niños sobre la IA en función de su comunidad, la diferencia de los impactos de la IA en los niños en los diferentes sectores, y la necesidad de ir más allá del reconocimiento de las necesidades de los usuarios, en la medida en que los niños pueden utilizar sistemas de los que no son los usuarios iniciales.

- + Es necesario empoderar a los jóvenes mediante la educación, la participación y el co-diseño de la IA. Mediante este empoderamiento, los niños pueden convertirse en defensores de sus propios derechos e influir en el desarrollo de la IA. Sin embargo, este empoderamiento no puede representar una sobrecarga de responsabilidades para los jóvenes, por lo que es necesario un enfoque de múltiples partes interesadas.
- + Es necesario educar a los niños, padres, profesores y cuidadores sobre los propósitos, los beneficios potenciales de la IA y sus procesos de desarrollo, entre otras cuestiones.
- + Debe haber una mayor claridad sobre cómo garantizar la aplicación del principio y quién es responsable de su aplicación.
- + El principio establece que los datos no deben utilizarse para causar daño, sin embargo, los datos también pueden tener impactos positivos en los niños y adolescentes.
- + Es necesario recoger datos sobre niños y adolescentes porque, de lo contrario, se pueden generar imprecisiones en los sistemas de IA que se les aplican. Por lo tanto, para equilibrar la recolección de datos con la seguridad de los jóvenes, deben utilizarse mecanismos como los sandboxes, la participación de los jóvenes o la evaluación continua de los sistemas.

La Mesa Redonda liderada por la CAF se orientó a recibir comentarios y retroalimentación sobre el Marco Ético, así como insumos sobre las mejores prácticas y experiencias de otros países. Los principales temas que se abordaron son la cocreación y la adopción de un enfoque multiactor, la intersección entre la ética de la inteligencia artificial y la ética de las tecnologías emergentes en general, las especificidades de Colombia para implementar el Marco Ético, la digitalización en el sector público y la intersección entre el Marco Ético y los marcos legales.

+ Principales recomendaciones de los participantes a considerar en la construcción del Marco Ético:

- +** Es necesario incluir un enfoque basado en el riesgo en la parte de los principios, ya que solamente se incluye en la parte de las herramientas.
- +** Debe haber diferencias entre la ética para el sector público y la ética para el sector privado ya que el sector público necesita un Marco Ético más estricto al ser responsable del bienestar de la gente.
- +** Debe haber un paso anterior al Marco Ético de IA que esté enfocado en los datos.
- +** Un Marco Ético se debe enfocar en las poblaciones diversas.
- +** Es necesario incluir a los ciudadanos ya que no habrá inteligencia artificial si los ciudadanos no son capaces de entender lo que se hace con la IA.
- +** Incluir el impacto positivo que puede tener la IA.
- +** Se debe considerar cómo deben incorporarse estos marcos éticos en el proceso de toma de decisiones en el sector público y el privado.
- +** Los principios deben encajar en una narrativa más amplia de la gobernanza de la IA. Tiene que haber una guía que asegure que la IA está alineada con lo que creemos.

La Mesa Redonda liderada por el BID mediante la iniciativa fAIrLACK se enfocó en cómo pasar de los principios a la práctica a partir de experiencias de distintos gobiernos y cómo estas últimas pueden ser un insumo para el caso colombiano. Los participantes fueron funcionarios de organizaciones internacionales y de gobiernos, profesores e investigadores de la academia y miembros de organizaciones promotoras de tecnologías emergentes. Los principales temas fueron la manera práctica de aterrizar los principios éticos, se definieron las audiencias a los que los principios van dirigidos, la relación del marco ético con la regulación, la regulación experimental y sandboxes, la protección de datos personales y las entidades dentro de los gobiernos encargadas de realizar el seguimiento a la implementación de los principios.

+

Principales recomendaciones de los participantes a considerar en la construcción del Marco Ético:

- + Es necesario entender a quién van dirigidos los principios ya que para los diferentes públicos (sector público y sector empresarial) el enfoque para la implementación de los principios debe ser diferenciado.
- + Es relevante tener en cuenta el contexto.
- + Establecer un cuerpo o entidad encargado de medir la implementación del Marco Ético.
- + Se debe asegurar la articulación entre el gobierno central y los gobiernos locales.
- + Se debe incluir los principios en todas las áreas posibles: academia, empresas privadas, ciudadanía.
- + Igualmente, es crítico entender la función, el alcance y el momento del tiempo de aplicación de las herramientas en un proyecto de IA.
- + La ciberseguridad debe ser incluida como un elemento fundamental de la seguridad y la estabilidad.

El webinar liderado por Co – Lab (Laboratorio de Innovación Educación Superior) tuvo como objetivo presentar el Marco Ético a los estudiantes ya que las instituciones de educación superior serán usuarios y creadores de soluciones de inteligencia artificial.

Los principales temas abordados fueron: la responsabilidad frente a la autonomía de algunos tipos de inteligencias artificiales, las herramientas para la materialización e implementación de los principios, la comercialización de las bases de datos y cómo convertir el país en líder del uso responsable de la IA.

+ Principales recomendaciones de los participantes a considerar en la construcción del Marco Ético:

- +** El documento debe ser considerado por diferentes grupos de la sociedad.
- +** Se debe mejorar la gobernanza y la ética en el AI Readiness Index.
- +** La academia debe ser líder en la ética de la inteligencia artificial.

La mesa liderada por el Ministerio de Educación Nacional de Colombia se centró en discutir el documento del Marco Ético desde la perspectiva del sector educativo. Esta mesa se inscribe en el contexto de la implementación del CONPES 3975, en el cual el ministerio de educación adquirió 7 frentes de trabajo, para los cuales el tema ético resulta fundamental.

Los participantes fueron principalmente directores de instituciones de educación superior, profesores e investigadores nacionales. Los principales temas abordados fueron: el enfoque interdisciplinario en el estudio de la inteligencia artificial más allá del componente técnico, la apropiación de la ética de inteligencia artificial en el currículo universitario, los desafíos de garantizar el cumplimiento de los principios del Marco Ético y el ente certificador encargado de realizar este seguimiento, la posibilidad de contar con tribunales de ética, y la transformación digital como vector de transformación de la cultura de las instituciones educativas.

+ Principales recomendaciones de los participantes a considerar en la construcción del Marco Ético:

- + Profundizar más en los conceptos, en vez de hablar de evaluación de algoritmos adentrarse más al concepto de verificación de algoritmos.
- + Se debe garantizar que los principios del Marco Ético se cumplan lo mejor posible.
- + Se propuso la academia o agencias especializadas, que no estén involucradas en el desarrollo del modelo que monitoreen y certifiquen que se están implementando los principios del Marco Ético.
- + Para la implementación del marco se deben incluir asignaturas de inteligencia artificial y de la ética de la inteligencia artificial en los currículos universitarios de ciencias sociales, de cara a que la inteligencia artificial no se limite al campo técnico. Para la implementación del marco también debe haber un cambio de cultura en las universidades.
- + El Marco Ético debe ser obligatorio porque de lo contrario se vuelve un saludo a la bandera, se deben encontrar mecanismos para profundizar su implementación.

En la mesa en la que participaron los representantes de los principales sectores económicos estos últimos presentaron sus aportes, inquietudes y consideraciones frente al Marco Ético.

Los principales temas abordados fueron: la protección de datos sensibles, la importancia de que el Marco Ético se traduzca en buenas prácticas y no en regulación todavía, la percepción de que se va a realizar una auditoría periódica de los algoritmos y surgieron inquietudes frente a la protección de la propiedad intelectual en la auditoría de los algoritmos, la autorregulación como alternativa, la evaluación y medición de los impactos de las normas, la colaboración entre los sectores público y privado en el análisis de los efectos de la inteligencia artificial.

+ Principales recomendaciones de los participantes a considerar en la construcción del Marco Ético:

- +** Promover que la adopción de los principios se traduzca en buenas prácticas y que estos no sean taxativos.
- +** Conformar un comité público-privado donde se analicen los efectos de la IA, y a partir de esto ir estudiando una alternativa de autorregulación.
- +** Enfatizar la seguridad de los datos sensibles en el principio de seguridad.
- +** Asegurar la alineación por parte del gobierno, entidades nacionales, sectores empresariales y educativos sobre el alcance de los principios.

Asimismo, se han incorporado los comentarios enviados por entidades del sector público, privado, organizaciones internacionales, la academia, ONG y la sociedad civil. Los principales comentarios recibidos fueron respecto a: la importancia de la contextualización frente a los riesgos de implementación de la IA y el nivel de necesidad de control humano, la relevancia de la sostenibilidad ambiental, la distinción entre la ética de la IA dentro de la función pública y en el sector privado, las métricas de desempeño, la privacidad y la responsabilidad.

+

Principales recomendaciones de los participantes a considerar en la construcción del Marco Ético:

- + Aclarar al usuario cuando está interactuando con un sistema y no con un ser humano.
- + Tratar de manera conjunta los principios de transparencia y explicación.
- + Adaptar los niveles de control humano a los grados de riesgo y al contexto.
- + Asegurar el respeto a la sostenibilidad ambiental en la implementación de IA.
- + Abordar la privacidad desde un marco más amplio de “privacy preserving” que engloba la privacidad diferencial, junto con otras propiedades.

Colombia, entendiendo los cambios abruptos que se están viviendo en materia de transformación digital, y en especial en la Inteligencia Artificial, e incorporando los insumos de las mesas técnicas, desarrolló el primer Marco Ético de América Latina.

La elaboración de este documento ha sido posible gracias a la cooperación técnica de la CAF. La Presidencia de la República agradece el apoyo técnico del BID (Banco Interamericano de Desarrollo) mediante la iniciativa fAIRLAC y de la universidad de Harvard mediante el Berkman Klein Center for Internet & Society, para la realización de este documento.



II

OBJETIVO Y ALCANCE DEL DOCUMENTO







OBJETIVO

El objetivo de este documento es brindar una “guía de soft law” de recomendaciones y sugerencias a las Entidades Públicas, para abordar la formulación y gestión de los proyectos que incluyan el uso de Inteligencia Artificial (IA). Esta guía reúne prácticas internacionales y no es de obligatorio cumplimiento por parte de las entidades, sino que se presenta como recomendaciones.



ALCANCE

Este documento sugiere la aplicación de principios éticos aplicables a los proyectos que hagan uso de Inteligencia Artificial, las herramientas para la implementación de los principios y algunas recomendaciones aplicables. Este documento está dirigido a los Líderes Estratégicos de TI, a los profesionales de las áreas de TI, a los gerentes de proyectos internos y externos de las entidades, a los profesionales del área de planeación, a la oficina de gestión de proyectos (PMO) y a los profesionales encargados de la implementación de la Política de Gobierno Digital en las Entidades Públicas del Estado Colombiano.



III

ADOPCIÓN DE PRINCIPIOS ÉTICOS PARA LA IA EN COLOMBIA







1o

TRANSPARENCIA Y EXPLICACIÓN:

Este principio debe ser entendido como la apertura para brindar información significativa y comprensible sobre el diseño, funcionamiento e impacto que tienen los sistemas de inteligencia artificial tanto para los desarrolladores y usuarios del sistema como para aquellos individuos que pueden verse afectados por sus decisiones y resultados. La información debe ser de fácil acceso y comprensible, en aras de promover una participación activa de los ciudadanos en la concepción, implementación y evaluación de los sistemas de IA. La información transparente sobre los sistemas de IA no debe comprometer la confidencialidad del modelo de gestión y de la innovación. Este principio es altamente contextual ya que la información y explicación deben adaptarse a un contexto y a una audiencia determinada. Por ejemplo, se deben tener en cuenta las perspectivas y características de las poblaciones diversas para garantizar su comprensión del funcionamiento de la IA y de sus implicaciones. La explicación debe ser todavía más prioritaria en aplicaciones de IA de alto riesgo.

- + **En la ética de los datos:** en la ética de los datos este principio tiene como efecto la necesidad de brindar información comprensible sobre la proveniencia de la información que se utiliza para el diseño de esta tecnología, las características de los datos y los fines para los cuales serán utilizados los datos, especialmente los datos personales. La transparencia y explicación hacen que las personas conozcan y comprendan la forma como se recolecta y procesa la información y los fines de hacerlo, en especial, cuando se procesan 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, y en los criterios utilizados para su clasificación y procesamiento.
- + **En la ética de los algoritmos:** significa que se provea información significativa y que se permita entender los objetivos iniciales que busca y que son propios del modelo, y los resultados esperados y también los obtenidos.

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 significativa y comprensible sobre los falsos positivos o negativos que ha elaborado un algoritmo y los indicadores de desempeño del algoritmo. Es claro que frente a los sistemas conocidos como de “*Black Box*” este principio puede entrar en tensión, dado que la capacidad de explicación puede llegar a ser ilimitada. 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, por qué son denominados *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 significativa, clara, y comprensible sobre los roles de los equipos y personas involucradas en el diseño, desarrollo e implementación 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. También en el marco de la implementación se debe informar a los usuarios en caso de relacionamiento con sistemas IA cuando no haya seres humanos detrás de los contenidos o respuestas que son generadas. Debe generarse también 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.



2

PRIVACIDAD:

La inteligencia artificial debe estar precedida de un respeto por la intimidad de las personas y su esfera privada que impide el uso de información que estos no hayan autorizado y el perfilamiento: comportamiento diferencial “con relación a diversos individuos en virtud de un criterio consistente en que tales individuos pertenezcan a grupos de población estadísticamente caracterizables de una cierta manera” (Peña, 2008) 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 assessments*). 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.





3 ◦

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 sistemas de inteligencia artificial que tengan 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. A lo anterior se suma que el control humano debe ser proporcional al nivel de riesgo, las aplicaciones de la tecnología que presenten una mayor probabilidad y severidad (Singapore Model Artificial Intelligence Framework) deben contar con mayor control humano frente a las de bajo riesgo.

- + **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 automatizada deben tener mecanismos para que los seres humanos intervengan y participen en cualquier momento.



4

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. La seguridad y confidencialidad de los datos personales y en especial de los datos sensibles son factores fundamentales para evitar afectaciones a la seguridad física y mental de los individuos.

- + **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. Las decisiones más trascendentales 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 salud o sector de seguridad nacional).
- + **En la ética de las prácticas:** se deben evitar aquellas prácticas que pongan en riesgo los sistemas de inteligencia artificial y los códigos de conducta deben generar parámetros para evitar aquellas actividades que pongan en peligro la integridad y seguridad física de las personas.



5

RESPONSABILIDAD:

Existe el deber de responder por los resultados que produzca un sistema de inteligencia artificial y las afectaciones que llegue a generar. Cada actor desempeña un papel dentro de la cadena algorítmica y no es posible limitar la responsabilidad a un participante específico ya que todos hacen parte de la misma cadena, 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.



6 ◊

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. 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. Los sistemas deben adoptar un enfoque de neutralidad de género y se debe garantizar que el parámetro de género no sea utilizado como factor de discriminación.

- + **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 tengan efectos discriminatorios.



7.

INCLUSIÓN:

Es la participación activa de poblaciones históricamente marginadas y de poblaciones diversas en el diseño, desarrollo, 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 las mujeres, 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. Es deseable desarrollar sistemas de inteligencia artificial que adopten un enfoque neutral desde el punto de vista de género y que no discriminen con base en el género de los individuos.



8

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 los niños, niñas y adolescentes, pero también de padres, profesores y cuidadores de esta tecnología y de sus implicaciones, que facilite la interacción que estos tengan con esta innovación y promueva una cultura de transformación digital. Con base en lo anterior se debe empoderar a los NNA para que participen activamente en el diseño de sistemas y de políticas públicas de IA, en una lógica de co-creación sin que esta participación genere una sobrecarga de responsabilidad que a estos no les corresponde. Por lo tanto, se debe adoptar un enfoque multiactor. Todo lo anterior se debe adaptar a los contextos de implementación de IA para los NNA y especialmente a los desafíos y oportunidades de las próximas generaciones.

- + **En la ética de los datos:** los datos de esta población únicamente pueden ser utilizados en actividades que se relacionen con su interés superior y que tengan un impacto exclusivamente positivo sobre los niños, niñas y adolescentes.
- + **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 en una lógica de co-creación 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.



9 ◊

BENEFICIO SOCIAL:

La inteligencia artificial tiene un potencial para generar efectos negativos pero también positivos los cuales resultan fundamentales ya que un escenario sin inteligencia artificial podría resultar perjudicial para el desarrollo y el bienestar. Por tanto, 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, el bienestar público la facilitación en la prestación de servicios públicos, la no afectación y protección del medio ambiente, el entretenimiento 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 políticas 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.



IV

HERRAMIENTAS PARA LA IMPLEMENTACIÓN DE LOS PRINCIPIOS PROPUESTOS



+



REGISTRO ÉTICO DE ALGORITMOS

Se busca que todas las entidades a las que se aplique este documento utilicen el registro ético de algoritmos (City of Amsterdam Algorithm Register Beta) que se ponga a disposición. El registro ético de algoritmos de inteligencia artificial consiste en una plataforma en la cual las entidades reportan la forma cómo están implementando la ética de la IA en sus proyectos. El Gobierno Nacional desarrollará la plataforma de registro ético de algoritmos para que las entidades reporten sus proyectos. El registro ético de algoritmos permite en primer lugar monitorear los avances de la implementación ética de la IA y en segundo lugar permite reforzar la participación ciudadana en la medida que les permite enviar comentarios o preguntas sobre las políticas, buenas prácticas o proyectos que se están implementando. Este registro es de acceso público e incluye un canal interactivo a través del cual los ciudadanos pueden formular dudas y comentarios acerca de la implementación ética de la inteligencia artificial. Las entidades deben reportar y actualizar periódicamente la siguiente información:

Proyectos:

- Descripción del proyecto
- Explicación de cómo la IA está siendo implementada en el proyecto
- Objetivo de Desarrollo Sostenible al cumplimiento del cual el proyecto contribuye
- Mecanismos para la implementación de los principios éticos
- Proceso de no discriminación
- Valoración de riesgos éticos y de seguridad, y medidas de mitigación para los identificados
- Aseguramiento del control humano (human over the loop)
- Entidad responsable
- Datos de contacto

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 política 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 indicaron 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

'LIMPIEZA' DE DATOS (DATA CLEANSING):

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:

- ▶ **Monitorear los errores:** llevar un registro y observar las tendencias sobre de 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).
- ▶ **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).
- ▶ **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).
- ▶ **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).

- ▶ **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).
- ▶ **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 susceptibles 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.



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).



4

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:

*We suggest that public officials consider **four key areas** to assess the effectiveness and legitimacy of an AI system's implementation:*

1. The process of dataset creation, e.g.:

- Who determines what data to collect?
- Who is included within the data?

2. The setup and design of AI tools, e.g.:

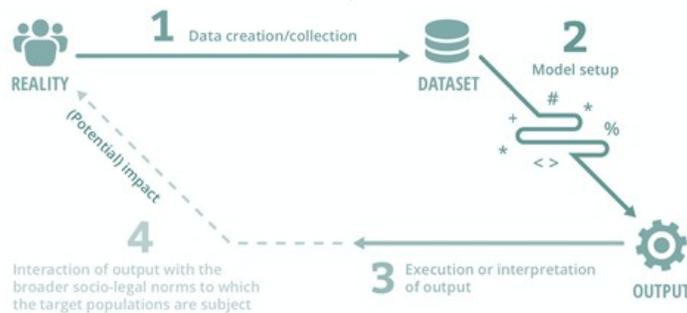
- What variables were included?
- Do they trigger risk of undue discrimination?
- Are outputs explainable? To whom? How?
- How do the outputs compare to human prediction or other non-AI systems?

3. The administrative protocols that surround the tool's output, e.g.:

- Does the tool inform human decisions, or execute policies directly?

4. Interaction with broader social and legal norms target populations are subject to, e.g.:

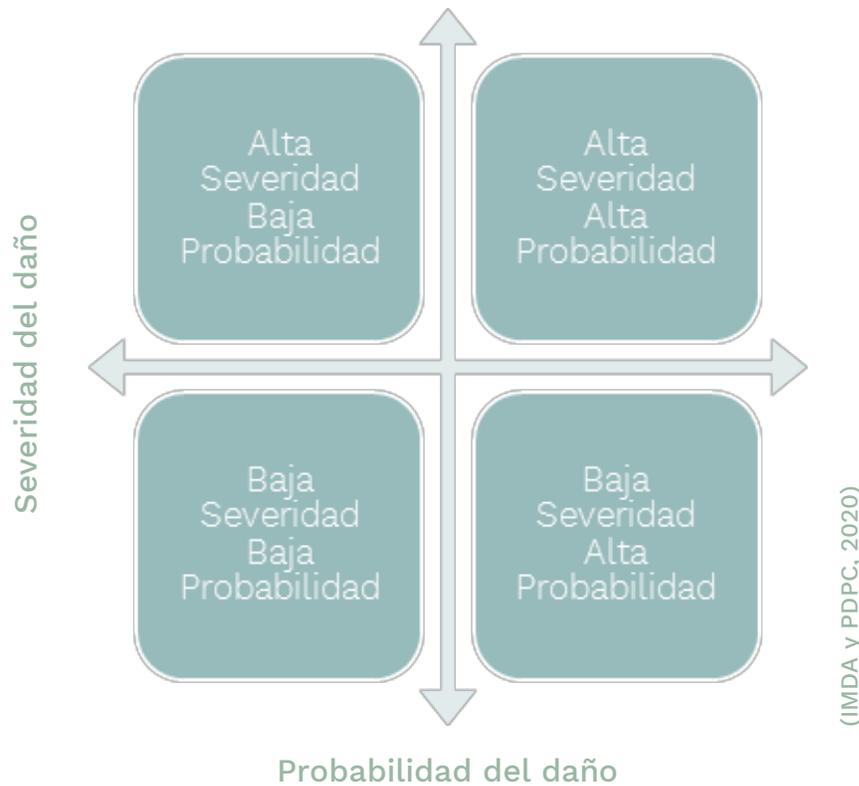
- Are there mechanisms of appeal for citizens who are impacted by decisions made by AI tools?
- What other safety-nets are available to those who are denied a service?
- How will the community treat a person who the AI classified in a certain way?



(World Wide Web Foundation, 2018)

- 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 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).



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 (Comisión Europea, 2020).

6 ●●●●●●●●●●

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 implemente un litmus test (código de conducta) en aras de 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, en esta óptica es relevante recordar a los empleados que piensen en los estándares de *big data* cuando planeen nuevos usos de los datos (ICO, 2017).

8 ●●●●●●●●●●

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 (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 deben estar entrenados para al menos estar alerta 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).



V

RECOMENDACIONES FINALES PARA EL SECTOR PÚBLICO





2

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 aprendizaje 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:

- Correr simulaciones extensivas (que contemplen múltiples escenarios de prueba para la búsqueda de medidas de restricción) durante la etapa de pruebas, para que las medidas apropiadas de restricción puedan programarse en el sistema (Leslie, 2019).

2

- Inspeccionar y monitorear continuamente el sistema, para que su comportamiento pueda ser mejor previsto y entendido (Leslie, 2019).
- Encontrar maneras de hacer que el sistema sea interpretado más fácilmente, para poder evaluar mejor sus decisiones (Leslie, 2019).
- Proveer mecanismos en el sistema que permitan a los humanos invalidar y apagar el sistema (Leslie, 2019).

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).

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 Inteligencia 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
- Centro Nacional de Consultoría. (2021, Enero 25). *CNC*. Retrieved from Centro Nacional de Consultoría: <https://www.centronacionaldeconsultoria.com/post/el-2020-fue-el-año-de-la-aceleración-de-la-transformación-digital-en-colombia-andi>
- 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
- DANE. (2021, Abril 15). *DANE*. Retrieved from DANE: <https://www.dane.gov.co/index.php/estadísticas-por-tema/servicios/encuesta-mensual-de-servicios-ems>
- 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/>

Referencias:

- 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 regulation*. 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>
- Navarro, J. G. (2021, Febrero 1). *Statista*. Retrieved from Statista: <https://www.statista.com/statistics/721751/latin-america-artificial-intelligence-market/>
- 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>

Referencias:

- 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
- Zaballos, A. G. (2021, Marzo 10). *BID*. Retrieved from iadb: <https://blogs.iadb.org/innovacion/es/inteligencia-artificial-como-catalizador-para-la-reactivacion-economica-y-el-crecimiento-productivo/>



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



El futuro
es de todos

Gobierno
de Colombia