
Policy Brief #2

Uso responsable de la inteligencia artificial en el sector público

Gobierno digital
e innovación pública

Uso responsable de la inteligencia artificial en el sector público

Tiempo de lectura: 20 minutos

Temas asociados: Inteligencia artificial – Ética – Gobernanza – Marco regulatorio – Sesgo – Uso responsable

Eje: gobierno digital e innovación pública

Nota de la Dirección de Transformación Digital de CAF -banco de desarrollo de América Latina

Esta nota está basada en el estudio “Uso responsable de la inteligencia artificial en el sector público” elaborado por María Isabel Vélez, Cristina Gómez Santamaría y Mariutsi Alexandra Osorio Sanabria, junto con Tibusay Sánchez Quintero (para el Capítulo 2), del Centro para la Cuarta Revolución Industrial de Colombia (C4IR.CO), afiliado al Foro Económico Mundial

Elaboración de la nota: Martha Rodríguez

Nota supervisada por: María Isabel Mejía, responsable de la Agenda de Gobierno Digital e Innovación Pública en CAF

Revisión y comentarios: Claudia Flores, Nathalie Gerbasi, María Isabel Mejía, Martha Rodríguez y Antonio Silveira

© 2022 Corporación Andina de Fomento

Las ideas y planteamientos contenidos en esta nota son de exclusiva responsabilidad de su autor y no comprometen la posición oficial de CAF.

Resumen

- Este *policy brief* es el segundo de una serie de documentos de política sobre el uso e impacto de la inteligencia artificial en el sector público. Estos documentos sintetizan los principales hallazgos, conclusiones y recomendaciones de **ExperienciaIA**, el reporte regional lanzado por CAF en septiembre de 2021, con el objetivo de sensibilizar sobre las oportunidades y los retos del uso estratégico de datos e inteligencia artificial (IA) en el sector público en la región.
- La IA es una tecnología poderosa, con enorme potencial para lograr Estados más eficaces y eficientes, que mejoren su capacidad para lograr impactos sociales, económicos y ambientales para el bienestar de los ciudadanos, siempre que se implemente de forma ética y estratégica.
- Los grandes desafíos de la IA para el sector público parten de la necesidad de que los sistemas de IA sean justos, eficientes y eficaces, con cuatro aspectos críticos en su diseño y operación: uso efectivo de los datos y la tecnología, capacidades humanas, cultura de lo público, y legitimidad y confianza.
- Si bien el desarrollo de la IA abre muchas posibilidades, existen riesgos potenciales que son necesarios considerar, comenzando por la posible discriminación de grupos e individuos, el uso inseguro o indebido de los datos, o la vulneración del derecho a la privacidad. Por tanto, uno de los retos que enfrentan los gobiernos es encontrar un equilibrio entre la explotación de una tecnología con la que pueden dar un salto en eficiencia y el establecimiento de límites a la misma para garantizar el bienestar social.
- Para avanzar en esta dirección, es necesario que el sector público preste especial atención a aspectos como la privacidad y confidencialidad de los datos, la transparencia y “explicabilidad” de los algoritmos, la inclusión, equidad o representatividad de los resultados obtenidos mediante IA, y la seguridad para impedir accesos no autorizados o manipulaciones de los datos.
- La ética de la IA es el conjunto de reglas y valores que se ajustan a lo que se considera correcto y aceptable para guiar su desarrollo y uso bajo conductas morales. A estos se suman principios y mecanismos que definen deberes y obligaciones básicos para el despliegue de sistemas de IA, que además, sean justos y seguros (Leslie, 2019).
- Aunque no hay una definición única de los principios éticos a aplicar para el despliegue de la IA, se distinguen dos elementos ampliamente adoptados: el reconocimiento de los daños que puede causar a los individuos y la sociedad sea por el abuso, el mal uso, problemas de diseño o efectos no deseados de la tecnología, y lo que se considera como deseable o beneficioso (reglas, valores, principios, técnicas) tanto a nivel individual como colectivo.

Hallazgos

1. Principales usos de la IA en el sector público

La adopción de sistemas de IA en el sector público puede generar beneficios en tres áreas que cubren buena parte de las responsabilidades y entidades de gobierno:

- Mejorar la formulación, ejecución y evaluación de las políticas públicas.
- Mejorar el diseño y la entrega de servicios a los ciudadanos y las empresas.
- Mejorar la gestión interna de las instituciones estatales.

Adicionalmente, esta tecnología puede direccionarse a temas específicos, como la salud, el transporte público, la defensa nacional, la educación o la administración de la justicia.

1.1 IA para mejorar la formulación, ejecución y evaluación de las políticas públicas

La IA ofrece oportunidades en las diferentes etapas del ciclo de las políticas públicas, que incluye cuatro etapas: identificación del problema e inclusión en la agenda política, diseño, implementación y evaluación.

4

Figura 1. Oportunidades que ofrece la IA en las diferentes etapas del ciclo de las políticas públicas



Fuente: Elaboración propia.

Identificación del problema e inclusión en la agenda

La IA puede ayudar en esta etapa de dos formas concretas: por una parte, identificando tendencias y anticipando situaciones emergentes que merecen la atención e intervención de las entidades públicas, tanto para el corto como

para el largo plazo, estar mejor preparados y pronosticar posibles resultados e impactos de las intervenciones. Por otra, captando y analizando los intereses y preocupaciones de los ciudadanos o diferentes grupos de interés para facilitar el alineamiento entre la agenda de gobierno y las necesidades e intereses de la población.

Diseño de políticas

La IA puede contribuir en el direccionamiento de las políticas hacia individuos, empresas o territorios en condiciones específicas o con necesidades más urgentes, para evitar el desperdicio de recursos y aumentar la probabilidad de obtener los resultados esperados. Así mismo, la IA permite comparar los posibles resultados de cada alternativa para la intervención pública, de acuerdo con alteraciones de diferentes variables, y también determinar el presupuesto y los recursos necesarios en cada caso, al analizar diferentes opciones, para elegir las más adecuadas para el contexto y los intereses específicos de las entidades y los Gobiernos.

Implementación

La IA puede impactar la implementación de políticas en la comunicación con diferentes audiencias dado que no todos los segmentos de la población perciben las políticas de la misma manera. Es posible adaptar el alcance, el tipo y las formas de interacción con la ayuda de esta tecnología, con el fin de aumentar las probabilidades de éxito de las intervenciones (Centre for Public Impact, 2017).

Igualmente permite modular la intensidad de las políticas en respuesta a cambios en el contexto en tiempo real, para considerar incluso situaciones inesperadas, como desastres naturales o epidemias, que pueden ser incorporados en los algoritmos de IA para ajustar o reorientar las políticas vigentes que lo ameriten (direccionamiento de ayudas y recursos, atención en centros de salud y alivios fiscales, entre otros).

Evaluación

El mayor impacto de la IA en la etapa de evaluación de las políticas en el corto plazo será la disminución de los tiempos necesarios para llevar a cabo sus actualizaciones o ajustes, proporcionando acceso a información valiosa en tiempo real para tomar decisiones sobre la necesidad de redireccionar, continuar o finalizar programas o proyectos. De esta forma, la evaluación, no se realiza luego de la implementación o al final del ciclo de la política, sino es un proceso que puede desarrollarse de manera continua (Valle-Cruz et al., 2020).

También permite considerar e integrar información de múltiples fuentes para realizar evaluaciones más integrales y completas, lo que puede ayudar a mejorar los procesos de retroalimentación y fortalecer la rendición de cuentas de las entidades responsables de las políticas.

1.2 La IA para mejorar el diseño y la entrega de servicios a los ciudadanos y las empresas

Como se ha señalado, la IA es útil para identificar intereses, preocupaciones y percepciones de diferentes actores, con el propósito de priorizar problemas a ser incluidos en la agenda de gobierno. Esta oportunidad puede favorecer un mejor entendimiento del comportamiento ciudadano o de ciertos grupos, más allá de la definición y priorización de los problemas públicos, hacia áreas que tienen que ver con la experiencia del acceso a servicios como el transporte, la salud, la educación, la seguridad o la justicia, o la manera en que grupos o ciudadanos se ven afectados por la actuación de las entidades públicas.

Esta información es fundamental para adaptar el diseño de los servicios a las maneras de pensar, sentir y actuar de las personas u organizaciones de acuerdo con sus realidades, lo que resulta muy útil, por ejemplo, para proveer servicios y asistencia a minorías o poblaciones específicas y diseñar servicios personalizados. También para hacer más eficientes las interacciones con los ciudadanos, con respuestas rápidas a cuestiones o solicitudes puntuales en sus versiones más básicas, o más sofisticadas para ofrecer orientación y respuestas de manera ágil posibilita mejorar los niveles de satisfacción de los ciudadanos con el desempeño de las entidades públicas.

Cabe destacar que es fundamental que los ciudadanos conozcan el uso que se está haciendo de los datos que ellos mismos producen en el diseño, desempeño y mejoramiento de los servicios públicos a los que tienen acceso. Solo así se podrá generar la confianza necesaria entre la población y las entidades públicas para que el intercambio de información se mantenga en el tiempo, facilitando el seguimiento del uso que se hace de los servicios públicos y, con ello, robusteciendo los sistemas de IA y sus procesos de aprendizaje, de manera que respondan mejor a situaciones específicas, necesidades y expectativas (van Ooijen *et al.*, 2019).

1.3 La IA para mejorar la gestión interna de las instituciones estatales

Las oportunidades ofrecidas por la IA se extienden también al funcionamiento y gestión de las entidades del Estado. No solo facilitan el cumplimiento de sus objetivos y responsabilidades, sino que permiten hacerlo incrementando los niveles de eficiencia y productividad. Con técnicas y herramientas que permiten la

predicción, la automatización, la optimización o el control, se logra un potencial enorme para mejorar el desempeño de cualquier organización, sea pública o privada.

Los sistemas de IA pueden apoyar la asignación y gestión de recursos financieros, ayudando a identificar y prevenir fraudes y el desvío o ineficiencias en la asignación y uso de dinero público, entre otros problemas. Así mismo, el procesamiento de solicitudes, requerimientos, análisis o decisiones puede hacerse de manera más rápida, ahorrando tiempo a las entidades y sus usuarios. En el caso de activos o infraestructura, es posible el mantenimiento preventivo, la corrección de fallas o la programación de su uso de acuerdo con la demanda mediante aplicaciones de IA, logrando una utilización más eficiente (van Ooijen *et al.*, 2019).

Por otro lado, la IA puede contribuir tanto a la generación y actualización de la regulación como al refuerzo de su cumplimiento. En el primer caso, la analítica de datos o las aplicaciones de aprendizaje automático ayudan a identificar vacíos o contradicciones normativas que deban resolverse, así como los aspectos más críticos en los que se debe enfocar la inspección o la aplicación de sanciones. La construcción de modelos predictivos permite detectar tempranamente fraudes u otras violaciones a la regulación. Recurriendo a estas aplicaciones, los esfuerzos de entidades regulatorias y de control pueden estar mejor dirigidos, priorizando aquellas situaciones donde es más pertinente y necesaria su intervención y dejando en un segundo plano actuaciones que no revisten gran relevancia o impacto.

Cabe resaltar además la posibilidad de agregar valor al trabajo público, pues por su capacidad de automatización, la IA puede asumir actividades y jui-

cios humanos repetitivos y rutinarios, facultando a los funcionarios que hacen este tipo de tareas para dedicar su tiempo, conocimientos y capacidades a explorar actividades de mayor valor y complejidad, donde la creatividad, el criterio, las habilidades emocionales y la perspectiva humana son necesarios.

Si bien las capacidades de la IA presentan un gran potencial para el sector público, ya sea para la optimización del proceso de la política, la provisión de servicios, la gestión de las entidades o el logro de objetivos en áreas de intervención específicas existe una alta incertidumbre respecto a los efectos y la evolución que pueden tener las aplicaciones de la tecnología en diferentes ámbitos. Hasta ahora, las experiencias con el desarrollo y uso de la IA, tanto a nivel público como privado, han llamado la atención sobre algunos aspectos que pueden conducir a resultados indeseados o problemáticos para los ciudadanos, las organizaciones, los Gobiernos y la sociedad en general. De llegar a materializarse, pueden deteriorar la confianza de los diferentes actores en la tecnología y, en consecuencia, minar su legitimidad para el uso público, asuntos que se abordan a continuación.

2. Los grandes desafíos de la IA para el sector público

El despliegue de la IA plantea a las entidades públicas un conjunto de desafíos derivados de la necesidad de que los sistemas de IA sean justos, eficientes y eficaces. Estos desafíos se refirieron a cuatro aspectos críticos del diseño y operación de sistemas de IA: el uso efectivo de los datos y la tecnología, los

recursos humanos, la cultura de lo público, y la legitimidad y confianza.

Uso efectivo de los datos y la tecnología

Para alcanzar un despliegue eficaz y ético de sistemas de IA que apoyen la toma de decisiones en el sector público, es indispensable acceder a datos precisos, que mantengan la privacidad y cumplan las normas éticas para su uso (Berryhill *et al.*, 2019). Por ello, las entidades públicas de orden nacional y local deben definir e implementar políticas y estrategias de gestión de los datos que permitan:

- Establecer lineamientos y estándares para orientar a los interesados en la implementación de programas de gestión de datos y de la tecnología de IA.
- Aumentar el uso y análisis de datos.
- Eliminar obstáculos a la disponibilidad, calidad, pertinencia, interoperabilidad, seguridad y privacidad de los datos.

Para asegurar la disponibilidad, los gobiernos deben contar con fuentes de datos internas y externas accesibles, oportunas y constantes para su aprovechamiento a través de la IA, así como disponer de diccionarios o catálogos de datos que faciliten su identificación, selección y descubrimiento (Berryhill *et al.*, 2019; Desouza, 2018). De igual modo, para compartir datos con otras entidades, es importante publicar conjuntos de datos relevantes en portales dedicados a ello y que se ajusten a las necesidades de los usuarios. Además, los gobiernos deben supervisar sus activos de datos y los métodos de recolección y evaluación de los datos existentes, de manera que puedan identificar eventuales ausencias de datos en las muestras a utilizar, las posibles variaciones en el tiempo y los sesgos que se puedan producir.

El éxito o fracaso en el análisis de datos a partir de los sistemas de IA depende en gran medida de la calidad de los datos. Para mantenerla es clave evaluar y mejorar los métodos de recopilación de los datos estructurados e identificar los que se necesitan, revisar si están disponibles y analizar si sus tipos y licencias de uso son adecuados para encontrar soluciones a la situación de interés y las necesidades de los interesados. Asimismo, asegurar que los responsables del procesamiento y análisis de datos utilicen fuentes adecuadas, así como mejorar el intercambio e interoperabilidad de los datos entre las diferentes áreas de las entidades públicas del orden nacional, regional y local (OCDE, 2019d).

Recursos humanos

El despliegue de la IA en el sector público requiere, como requisito fundamental, la presencia de empleados que entiendan la tecnología y el potencial que esta ofrece, no solo a nivel técnico, sino también a nivel operativo y directivo.

La capacitación de los empleados públicos debe incluir la actualización y perfeccionamiento de ciertas habilidades, que serán necesarias para desempeñar las nuevas funciones que emerjan a medida que se adopte la IA. Algunas de ellas podrán exigir una interacción directa con los sistemas desarrollados, la interpretación de sus resultados o su monitoreo. Al mismo tiempo, la automatización implicará la reubicación de un número significativo de trabajadores, para lo cual es importante una articulación entre el sector público, las empresas y las instituciones, de manera que se facilite esa reacomodación de la fuerza laboral.

Cultura de lo público

En el sector público se presentan barreras para la adopción de la IA con pocos incentivos a la innovación. Los empleados públicos no suelen ser estimulados a tomar riesgos o experimentar y, en el caso de los líderes organizacionales, la poca disposición a correr riesgos o la ausencia de una mentalidad innovadora puede perjudicar su compromiso con la adopción de nuevas tec-

nologías transformadoras como la IA y generar un ambiente de poca apertura al cambio (Ubaldo *et al.*, 2019).

Resistencia a su incorporación en el proceso público y los cambios que pueda generar, debida al desconocimiento del alcance y las posibilidades que ofrece la tecnología, sumado al temor que para muchos funcionarios despierta la automatización de algunas de las tareas que realizan.

Características de los procesos de adquisición o compra pública. Muchos desarrolladores consideran los algoritmos como activos de propiedad intelectual, lo cual impide que las entidades puedan eventualmente adaptarlos o actualizarlos para mantenerlos relevantes y pertinentes (WEF, 2020a). Adicionalmente, los procesos de contratación son lentos, complejos y muy específicos, lo que dificulta la provisión de soluciones de manera oportuna y entorpecen en ocasiones la participación de proveedores, particularmente empresas pequeñas o emergentes (*startups*).

Legitimidad y confianza

Uno de los mayores retos para el sector público, si no el principal, es lograr que la ciudadanía acepte, confíe y respalde el uso de la IA por parte de los gobiernos. Para conseguirlo estos tienen que ofrecer garantías a los ciudadanos ante usos indebidos o posibles perjuicios derivados del uso de algoritmos y, por otro lado, deben aumentar su propia eficiencia, optimizar el ciclo de las políticas públicas y la atención a los ciudadanos usando justamente algoritmos. Esto es, «governar algoritmos mientras se gobierna con algoritmos» (Kuziemski y Misuraca, 2020).

Los gobiernos deben adquirir capacidades para generar respuestas y mecanismos que minimicen los riesgos y preocupaciones que se derivan del uso de la IA, tanto en el ámbito público como en el privado, alrededor de asuntos como la privacidad y la transparencia. El uso no autorizado de datos personales o su divulgación constituyen una amenaza para la protección del derecho a la privacidad y generan cuestionamientos sobre el alcance que tiene la tecnología y cómo puede ser usada en contra de grupos o individuos.

Asimismo, la complejidad de ciertos algoritmos, cuándo se usan y para qué, son asuntos que exigen claridad por parte del Gobierno de manera que su actuación y decisiones puedan justificarse y entenderse.

Al mismo tiempo, situaciones de discriminación o exclusión, entre otros efectos indeseables causados por decisiones basadas en IA, muchas veces sobre individuos o grupos que no cuentan con las herramientas necesarias para pronunciarse ante tratos injustos, son factores que debilitan la confianza ciudadana en la tecnología. Con frecuencia es la presencia de sesgos lo que conduce a dichas situaciones discriminatorias.

3. El sesgo en la inteligencia artificial

Los sesgos en los sistemas de IA se refieren a errores, inexactitudes o anomalías que se presentan en los algoritmos con los que opera, pudiendo tener impactos no deseados en individuos y grupos, principalmente en términos de discriminación o exclusión; lo que determina si este es o no es justo. La presencia de sesgos representa un serio riesgo para las entidades públicas.

Los sesgos en los sistemas de IA pueden darse principalmente por: i) suposiciones o prejuicios que se introducen durante el diseño y desarrollo del algoritmo, el cual es realizado por humanos que tienen sus propios sesgos cognitivos y sociales, los cuales terminan siendo integrados en los parámetros que se definen para el sistema; y/o ii) suposiciones o prejuicios sociales o históricos, introducidos durante la recolección o etiquetado de los datos que se emplean para entrenar el sistema, o por su falta de representatividad, es decir, la ausencia de datos.

Además de mantener, reafirmar o amplificar disparidades o situaciones de exclusión existentes en el mundo real, los sistemas de IA sesgados pueden, por ejemplo, afectar la obtención de subsidios y otras ayudas sociales o el acceso a servicios públicos, como el de educación, salud o la administración de la justicia (Spielkamp, 2017).

A pesar de la importancia de prevenir y minimizar los sesgos en los sistemas de IA, no existe todavía una fórmula exacta para hacerlo, pero es posible realizar acciones concretas que ayuden a prevenirlos y mitigar su efecto. Es necesario invertir tiempo y recursos para auditar los datos, especialmente cuando información sensible, como la edad, el género o la raza, forman parte del conjunto de datos empleados para entrenar el sistema de IA (Clausen, 2020; McKenna, s. f.). Como los sesgos, que pueden afectar de manera significativa las decisiones tomadas por un algoritmo, pueden presentarse en varias etapas del proceso, la auditoría de los datos no es suficiente. Para implementar en la práctica modelos no sesgados, es necesario realizar un monitoreo constante durante todo el ciclo de vida del sistema y contar con un modelo de gobernanza claro (Clausen, 2020).

Para tratar de que no se introduzcan sesgos y prejuicios propios del desarrollador en el diseño y construcción de los algoritmos, es recomendable contar con grupos multidisciplinarios y heterogéneos que permitan minimizar este riesgo (Shin, 2020; McKenna, s. f.). De igual forma, someter los algoritmos a la evaluación de distintos grupos antes de una etapa de producción puede ayudar a identificar los sesgos potenciales (Shin, 2020).

4. Ética de la IA y de los datos

El concepto de ética hace referencia al estudio de la moralidad, que se entiende como un sistema de reglas y valores que guían la conducta humana, junto con los principios para evaluar esas reglas. En consecuencia, cuando se dice que un comportamiento es ético, no estamos diciendo necesariamente que sea un «buen» comportamiento, sino que se ajusta a ciertos valores (WEF, 2019a).

Aplicada a la IA, la ética se entiende como el conjunto de reglas y valores que se ajustan a lo que se considera correcto y aceptable para guiar su desarrollo y uso bajo conductas morales. A esos valores y reglas se suman principios y otros mecanismos que definen deberes y obligaciones básicos para el despliegue de sistemas de IA, que, además de éticos, sean justos y seguros (Leslie, 2019).

Para definir la ética de la IA es fundamental, por un lado, el reconocimiento de los daños que puede causar a los individuos y a la sociedad ya sea por el abuso, el mal uso, problemas de diseño o efectos no deseados de la tecnología, y por otro lado, lo que se considera como deseable o beneficioso (reglas, valores, principios, técnicas) tanto a nivel individual como colectivo, en oposición a dichos daños.

Las diferentes posiciones respecto a lo que se considera deseable o beneficioso para los individuos y la sociedad frente a los posibles perjuicios de la IA pueden conciliarse mediante la creación de marcos éticos para el diseño y despliegue de esta tecnología.

Los principios adoptados más frecuentemente por varias instituciones internacionales y en

América Latina¹ incluyen la privacidad, responsabilidad, seguridad y protección, transparencia y «explicabilidad», equidad y no discriminación, control humano de la tecnología, promoción de las capacidades, valores y derechos humanos, y autonomía humana.

Una vez establecidos los principios éticos para el despliegue de la IA, el paso siguiente es generar alternativas para llevarlos a la práctica, lo cual se logra combinando medidas técnicas y no técnicas. Las primeras incluyen el cumplimiento de la ética «desde el diseño», la cual puede enfocarse de diferentes maneras (WEF, 2019a).

Otras alternativas técnicas para el cumplimiento de los principios éticos incluyen contar con métodos de explicación, como los que se vienen investigando en la «IA explicable»; la realización de ensayos y validaciones constantes para hacer un seguimiento minucioso de la estabilidad, solidez y funcionamiento del modelo durante su ciclo de vida; y el empleo de métricas e indicadores de calidad de servicio previamente definidos, que permitan evaluar el desempeño del sistema, tales como rendimiento, usabilidad, fiabilidad, seguridad y facilidad de mantenimiento (Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial, 2019).

Las alternativas no técnicas contemplan, entre otras, la revisión y adaptación de la normatividad vigente, la expedición de códigos de conducta generales y sectoriales, la educación y concientización de quienes diseñan, implementan, usan y se ven afectados por la IA, certificaciones, marcos de gobernanza y, por supuesto, regulación.

¹ Como OCDE, la Comisión Europea, el Grupo independiente de expertos de alto nivel sobre IA, IA-Latam, IEEE y la Iniciativa global sobre ética de sistemas autónomos e inteligentes

5. Gobernanza de la IA en el sector público

La innovación tecnológica tiene el potencial de aumentar la productividad, el crecimiento económico y, en general, el bienestar de las personas. Sin embargo, como se ha mencionado, puede tener consecuencias negativas que obligan a cuestionar si las regulaciones actuales en los muchos ámbitos de aplicación son adecuadas. El reto de las instituciones de gobierno es definir políticas públicas que prevengan, mitiguen y corrijan los posibles efectos negativos de las nuevas tecnologías sin restringir las posibilidades de emprendimiento que ofrecen, ni impedir sus potenciales beneficios (OCDE, 2018a).

La gobernanza tecnológica se define como el «proceso de ejercer la autoridad política, económica y administrativa en el desarrollo, difusión y operación de tecnología en las sociedades» (OCDE, 2018a) e implica la orientación y definición de objetivos y la elección de los mecanismos para alcanzarlos, la regulación de su funcionamiento y la verificación de sus resultados, entre otros aspectos (Stirling *et al.*, 2019). En el caso particular de la IA, los marcos de gobernanza deben ser objeto de revisión, ajuste y actualización permanente a medida que la tecnología y sus aplicaciones evolucionan, tanto en el ámbito público como en el privado.

El Foro Económico Mundial (WEF, por sus siglas en inglés), por ejemplo, propone, en el marco de la Cuarta Revolución Industrial (4RI), la implementación de una gobernanza ágil, entendida como la disposición para navegar rápidamente el cambio generado por las tecnologías, aceptarlo y aprender de él mientras se contribuye al valor real o percibido por el

usuario final. También hace referencia al diseño adaptativo de políticas, centrado en el ser humano, inclusivo y sostenible, que reconoce que el desarrollo de políticas no se limita a los Gobiernos, sino que es un esfuerzo de múltiples actores (WEF, 2018a).

Una gobernanza responsable de la IA debe contemplar un marco de gobernanza que sea anticipatorio a la exploración de la tecnología, sus propósitos y posibles trayectorias de desarrollo; inclusivo, para una vinculación activa y el diálogo entre diversos públicos y actores desde las primeras fases de los desarrollos tecnológicos; adaptable para cambiar visiones, comportamientos y estructuras organizacionales con el fin de dar respuesta a nuevas circunstancias, perspectivas, conocimientos y normas; y con propósito para direccionar la tecnología en razón a lo que socialmente se considera deseable². Este marco sirve para orientar no solo la actuación del sector público, sino también el despliegue general de la tecnología en la sociedad.

La creación de esos marcos de gobernanza implica desplegar un amplio conjunto de mecanismos, herramientas y prácticas que, desde distintos frentes (técnico, de política pública, legal, de relación con actores interesados u otros), apunten al uso de la tecnología en favor del interés común. Los principios pueden divulgarse mediante guías y herramientas para la auditoría de algoritmos y la evaluación de impactos y riesgos, entre otros.

Para la construcción de la gobernanza importantes tres aspectos específicos: la gobernanza de los datos, la evaluación de riesgos e impactos y las estructuras y medidas para la operación de sistemas de IA.

² Estos atributos de la gobernanza se inspiran en el enfoque de la investigación e innovación responsables (RRI, por sus siglas en inglés), que busca anticipar y evaluar las posibles implicaciones y expectativas sociales respecto a la investigación y la innovación, con el propósito de hacerlas éticamente aceptables, sostenibles y socialmente deseables. Para ello, señala que los procesos de investigación e innovación deben ser diversos e inclusivos, anticipativos y reflexivos, abiertos y transparentes, con capacidad de respuesta y adaptativos. Para ampliar, consultar la web siguiente: <https://www.rri-tools.eu/es/about-rri#wha>

La gobernanza de los datos

Es la manera de ejercer control sobre la calidad de los datos, el cumplimiento de los requerimientos legales y éticos relacionados con los datos y su utilización para la creación de valor público, garantizando resultados confiables. Es especialmente relevante pues los datos que nutren la IA provienen de diferentes fuentes, lo que incrementa la dependencia entre organizaciones y dificulta la asignación de responsabilidades. Además, errores en los datos pueden traducirse en decisiones sesgadas o ilegales, altos riesgos financieros y crisis políticas, entre otros, lo que tiene serias implicaciones para

las organizaciones involucradas, los ciudadanos, las empresas y la sociedad en general. Para lograr su propósito, un buen esquema de gobernanza de datos para la IA requiere, entre otros, mecanismos de protección de datos personales, seguridad, no discriminación e igualdad en el trato, cubriendo toda la cadena de valor del dato y abordando las implicaciones técnicas, institucionales y sociales del intercambio de datos (Janssen *et al.*, 2020).

El Cuadro 1 presenta consideraciones relevantes para estructurar el marco de gobernanza de datos, especialmente aquellos usados por los sistemas de IA.

Cuadro 1 Resumen de principios para la gobernanza de datos

PRINCIPIO	DESCRIPCIÓN
Evaluar la calidad y el sesgo de los datos	Cuando los datos son usados en sistemas algorítmicos, es necesario evaluar su calidad y sesgos.
Detectar patrones cambiantes	Cuando los resultados de los algoritmos cambian, se debe verificar su validez e investigar las razones de tales cambios.
Información requerida	Minimizar la cantidad de datos que se comparten (solo lo necesario); por ejemplo, respuestas a preguntas en lugar de conjuntos de datos completos.
Recompensa por errores	Las recompensas podrían usarse para alentar a las personas a detectar errores y reportarlos.
Informar al compartir	Cuando los Gobiernos comparten datos sobre una persona u organización, estos deben ser informados para garantizar la transparencia y evitar el uso indebido.
Separación de datos	Separar los datos personales de los no personales y los datos sensibles de los no sensibles.
Control ciudadano de los datos	Empoderar a las entidades públicas, los ciudadanos y las organizaciones para que tengan el control y verifiquen la precisión de sus datos.
Recopilación de datos en la fuente	Para garantizar su exactitud y saber cómo se recaban.
Minimizar la autorización para acceder a los datos	Si una de las partes no necesita datos, no se le debe otorgar acceso.
Almacenamiento distribuido de datos	Los sistemas distribuidos son menos vulnerables y evitan combinar datos fácilmente sin autorización.
Administradores de datos	Asignar administradores de datos para formalizar la rendición de cuentas por la gestión de los recursos de información mientras se adhiere a la separación de preocupaciones.
Separaciones de preocupaciones	Las responsabilidades respecto a los datos deben distribuirse de tal manera que una única persona no pueda hacer un mal uso o abusar de los mismos.
Utilidad	Los datos deben reconocerse como un activo valioso que puede ser utilizado para la IA

Fuente: Janssen *et al.* (2020)

Evaluación de riesgos e impactos

La evaluación de los posibles impactos que tendría el uso de un algoritmo de IA sobre individuos, empresas y comunidades, asociado a la clara identificación de los riesgos involucrados, es un requisito indispensable para determinar las acciones necesarias que garantizarán el cumplimiento de los principios éticos de la IA, entre ellas el nivel de participación humana. Estos análisis deben ser monitoreados durante el ciclo de vida del sistema de IA y ajustados cuando sea necesario.

Una primera evaluación debe contemplar las necesidades de los usuarios, las comunidades afectadas, los potenciales riesgos y sesgos en el sistema, así como escenarios de consecuencias no deseadas. Para estos últimos, es útil realizar el análisis del impacto específico sobre la protección de datos y la igualdad, considerando cómo el uso del sistema interactúa con mecanismos de supervisión, revisión y otras salvaguardas (WEF, 2020a).

En cuanto a la evaluación de riesgos, a continuación se presentan los aspectos clave a considerar, de acuerdo con el WEF (2020a).

Cuadro 2. Aspectos clave de la evaluación de riesgos

DATOS	Sensibilidad. Cuanto más confidenciales los datos, mayores puntos de chequeo deben incluirse. Es necesario considerar si los datos pueden identificarse o revelar información personal.
	Calidad. Ante dudas sobre la calidad de los datos, se recomienda incorporar garantías adicionales para evitar sesgos y reducir el riesgo del proyecto. Para asegurar la representatividad es posible que se requieran medidas específicas. Se deben considerar, además, sesgos sociales que podrían reflejarse en los datos.
	Consentimiento. Es necesario tener autorización de uso de los datos para el área de aplicación específica del sistema de IA y hay que asegurar que ese uso no sea inferido a partir de otros consentimientos.
ÁREA DE USO	Escrutinio público. Un proyecto en sectores de servicios con especial sensibilidad, como los de salud, asistencia social, empleo, finanzas, justicia e inmigración, entre otros, exigen mayores consideraciones y precauciones.
IMPACTO SOCIOECONÓMICO	Actores involucrados. Cuanto mayor sea el impacto potencial en individuos, empresas o comunidades, más importantes serán las consideraciones éticas y el análisis de la solución de IA.
	Alcance del impacto. Es importante tener en cuenta cuántas personas serán impactadas y qué tan alto y probable es dicho impacto. El riesgo se incrementa cuando son grupos vulnerables.
CONSECUENCIAS FINANCIERAS PARA LA AGENCIA E INDIVIDUOS	Alcance del impacto financiero. Cuantos mayores sean las posibles consecuencias financieras, más se deben abordar todas las áreas relacionadas con las consideraciones específicas de la IA.
	Tipos de impactos financieros. Pueden incluir aspectos monetarios, acceso al crédito, oportunidades económicas, escolaridad o formación, seguros y certificaciones.
IMPACTO DEL SISTEMA DE IA EN SUS PROCESOS, EMPLEADOS Y ACTIVIDADES PRINCIPALES	Impacto en las actividades centrales. A mayor dependencia tecnológica, mayor el riesgo. Además de los riesgos técnicos, deben mitigarse los riesgos reputacionales.
	Impacto en las funciones de la agencia. Considerar si se reemplaza una función o solo se mejoran o aumentan las actuales.
	Pérdida de empleos. Mayor automatización puede conducir a mayor pérdida de empleos, aumentando los riesgos y sensibilidad respecto al despliegue de la IA.
	Intervenciones humanas. Pocos controles y contrapesos generan mayor riesgo. Es necesario que las soluciones que surgen como resultado de la IA se puedan explicar e interpretar.

Fuente: Basado en WEF (2020a).

Lo más importante de las evaluaciones de riesgos e impactos es que conduzcan al diseño de estrategias para atenderlos y gestionarlos de la mejor manera (por ejemplo,

intervención humana, revisión de pares, documentación y explicación del algoritmo, pruebas, consultas a la ciudadanía). Los funcionarios responsables de la toma de deci-

siones deben conocer particularmente los riesgos y las medidas para mitigarlos y, si es el caso, dar fundamento a la continuidad o cancelación de un proyecto (WEF, 2020a).

Estructuras y medidas para la operación de sistemas de IA

Las estructuras de gobernanza son indispensables para que las entidades públicas puedan supervisar y controlar lo que ocurre durante el ciclo de vida de los sistemas de IA. Pueden ser creadas para atender específicamente las cuestiones relacionadas con la tecnología o resultar de adaptaciones o actualizaciones de estructuras existentes, de acuerdo con la situación particular de cada entidad.

la poca proactividad. Esta situación está llevando a los Estados a explorar nuevos enfoques regulatorios que sean más ágiles y colaborativos, que implican, entre otros factores, una buena combinación de ambos tipos de instrumentos (vinculantes y no vinculantes), interdisciplinariedad y participación de diferentes actores.

Los principales desafíos que encuentra la regulación de IA en la actualidad incluyen la evaluación y utilización de la regulación vigente, siempre y cuando permiten a muchos países manejar ciertos aspectos de la IA, por ejemplo, los relacionados con la privacidad. Así mismo, regular en la medida justa, las tensiones sobre la forma de regular, con diferentes visiones sobre las áreas y actividades que deben regularse varían ampliamente entre regiones, al igual que la falta de consenso sobre qué dirección seguir para avanzar, dados los cambios en la tecnología y la rápida evolución de la IA (WEF, 2020c).

Se destacan dos tipos de alternativas para abordar la regulación:

- **Leyes y normas (*hard law*).** El desarrollo de una regulación específica se justifica en la medida en que los riesgos que plantea esta tecnología son nuevos o no están contemplados por la regulación actual. Sin embargo, es importante considerar hasta qué punto las leyes y normas (*hard law*) existentes pueden regularlos adecuadamente o el tipo de ajustes que serían necesarios para que eso ocurra. A pesar del creciente número de países que ha decretado una estrategia nacional para la IA, la mayoría de los Estados han sido cautelosos en lo que respecta a la creación de instrumentos vinculantes específicos. Los principales avances en esa dirección se han dado en relación con las leyes de uso y protección de datos personales, las cuales tienen una incidencia directa sobre gran parte de los desarrollos de la IA.
- **Instrumentos no vinculantes.** Hasta ahora, las acciones tendientes a regular la IA se han concentrado en el desarrollo de instrumentos no vinculantes (*soft law*). Entre ellas

6. Un ecosistema de confianza: marco regulatorio para la IA

La regulación se refiere al conjunto de reglas, normas y sanciones formales e informales que buscan moldear el comportamiento de las personas para lograr un objetivo o meta de política. Su objetivo es brindar certeza y gestionar los riesgos, al tiempo que permite que los beneficios se distribuyan de manera equitativa (WEF, 2020c).

Los sistemas regulatorios tradicionales comprenden instrumentos legalmente vinculantes —como leyes, decretos, resoluciones— e instrumentos no vinculantes —como guías y códigos de práctica, autorregulación, estándares y certificaciones, entre otros—. Estos instrumentos buscan influenciar el comportamiento de una forma menos directa y se clasifican dentro de las medidas no técnicas para implementar los principios éticos de la IA. En el caso de las tecnologías emergentes, la complejidad técnica y la velocidad de los cambios exponen diferentes limitaciones de estos sistemas regulatorios, entre las que figuran la poca flexibilidad y

figuran guías³, marcos éticos y herramientas para el análisis de riesgos e impactos, como los mencionados en secciones previas y que pueden también considerarse instrumentos de autorregulación. Una alternativa interesante, que viene surgiendo para replantear los modelos de regulación actuales ante la incertidumbre y expectativa generada por el avance de la IA y sus posibles impactos en la sociedad, es la creación de ambientes controlados donde puedan ponerse a prueba sistemas de IA. Estos espacios de experimentación facilitan la prueba de innovaciones a pequeña escala (productos, servicios, modelos de negocio) en un entorno controlado, similar al del mercado, suspendiendo temporalmente reglas, disposiciones o requisitos obligatorios a cambio de que se incorporen salvaguardas adecuadas para aislar al mercado de los riesgos de la innovación (Zetzsche Buckley *et al.*, 2017; FCA, 2017).

Adicionalmente existen diversos modelos de agencias regulatorias, que pueden adoptar diversas formas, responsabilidades y atribuciones. Por otro lado, pueden estar enfocadas en una tecnología o grupo de tecnologías, una industria, un problema o un objetivo de política. En cual-

quier caso, resulta clave definir su alcance, con opciones diversas, que incluyen la creación de una única agencia (transversal) a nivel nacional, que opere en todos los campos de uso y disciplinas, incluyendo el sector público y el privado; agencias especializadas (verticales), encargadas de la supervisión de los algoritmos en sus diferentes campos de aplicación, siguiendo el modelo usado para determinadas industrias, como la farmacéutica o de alimentos, la educación o el sistema financiero; crear un órgano específico como custodio y encargado de la regulación de la IA, que monitoree y solicite explicaciones sobre la forma en que los algoritmos toman decisiones u otros organismos especializados (Abdala *et al.*, 2019).

Avanzar en la construcción de marcos regulatorios para la IA es crucial para definir una regulación centrada en la ética, la seguridad, la justicia y la transparencia, que fortalezca la confianza de la ciudadanía en la tecnología y su uso por parte del Estado. Dado que este proceso se encuentra en una etapa inicial en la mayoría de los países, hay una oportunidad para que se discuta y comparta con todos los actores involucrados, además de presentar los avances para revisiones y control público frecuente para mantener su relevancia, aplicabilidad y efectividad.

³ Un ejemplo es la «Guía para la construcción y el uso de la IA en el sector público», publicada en el año 2019 por el Servicio Digital del Gobierno y la Oficina de Inteligencia Artificial del Reino Unido (GDS y OAI, 2019). Este documento ofrece orientaciones en cinco frentes, que van desde el entendimiento de la tecnología y la evaluación de su pertinencia, pasando por su planeación e implementación hasta la comprensión de aspectos éticos y de seguridad. Cada uno de estos frentes plantea definiciones, acciones concretas y ejemplos en un lenguaje concreto y sencillo para los encargados de tomar decisiones relacionadas con la IA en el sector público.

Bibliografía

Abdala, M., Eussler, S. L. y Soubie, S. (2019). «La política de la inteligencia artificial: sus usos en el sector público y sus implicaciones regulatorias». Documento de trabajo n.o 185. CIPPEC. <https://www.cippec.org/wp-content/uploads/2019/10/185-DT-Abdala-Lacroix-y-Soubie-La-pol%C3%ADtica-de-la-Inteligencia-Artificial-octubre-2019.pdf>.

Berryhill, J., Heang, K. K., Clogher, R. y McBride, K. (2019). «Hello, world: artificial intelligence and its use in the public sector». *OECD Working Papers on Public Governance*, n.o 36. <https://doi.org/10.1787/726fd39d-en>.

Clausen, K. G. (2020, marzo 11). «How to tackle bias in AI». 2021.AI [en línea]. <https://2021.ai/how-to-tackle-bias-in-ai/>.

Desouza K. (2018). Delivering artificial intelligence in government: challenges and opportunities. IBM Center for Business of Government. <http://www.businessofgovernment.org/sites/default/files/Delivering%20Artificial%20Intelligence%20in%20Government.pdf>.

FCA (2017). Regulatory sandbox lessons learned report. Financial Conduct Authority. <https://www.fca.org.uk/publication/research-and-data/regulatory-sandbox-lessons-learned-report.pdf>.

GDS y OAI (2019). *A guide to using AI in the public sector*. Government Digital Service y Office for Artificial Intelligence. Actualizado el 18 de octubre de 2021. <https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector>.

Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial (2019). *Directrices técnicas para una IA fiable*. Comisión Europea. <https://ialab.com.ar/wp-content/uploads/2020/06/Grupo-independiente-de-expertos-de-alto-nivel-sobre-IA-creado-por-la-Comisio%C3%81n-Europea.pdf>.

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S. y Janowski, T. (2020). «Data governance. Organizing data for trustworthy artificial intelligence». *Government Information Quarterly*, 37(3), 101493. <https://doi.org/10.1016/j.giq.2020.101493>.

Kuziemski, M. y Misuraca, G. (2020). «AI governance in the public sector: three tales from the frontiers of automated decision-making in democratic settings». *Telecommunications Policy*, 44(6), 101976. <https://doi.org/10.1016/j.telpol.2020.101976>.

Leslie, D. (2019). *Understanding artificial intelligence ethics and safety. A guide for the responsible design and implementation of AI systems in the public sector*. The Alan Turing Institute. <https://doi.org/10.5281/zenodo.3240529>.

McKenna, M. (s. f.). «Machines and trust: how to mitigate AI bias». *Toptal Engineering* [en línea]. Blog. <https://www.toptal.com/artificial-intelligence/mitigating-ai-bias> (consulta realizada el 3 de marzo de 2021).

OCDE (2018a). *OECD science, technology and innovation outlook 2018: adapting to technological and societal disruption*. OCDE. https://doi.org/10.1787/sti_in_outlook-2018-en.

OCDE (2019a). *Data in the digital age*. Paris: OECD Publishing. <https://doi.org/10.1787/99b9ba70-en>.

OCDE (2019b). «Using digital technologies to improve the design and enforcement of public policies». *OECD Digital Economy Papers*, n.o 274). <https://dx.doi.org/10.1787/99b9ba70-en>.

OECD. (2019b). *The Pathway to Becoming a Data-Driven Public Sector*. OECD. <https://doi.org/10.1787/059814a7-en>.

OCDE (2019c). *Artificial intelligence in society*. Paris: OECD Publishing. <https://doi.org/10.1787/eedfee77-en>.

OCDE (2019d). «Government at a glance 2019 edition. Open government data». OECD.Stats [base de datos]. <https://stats.oecd.org/Index.aspx?DataSetCode=GOV> (consulta realizada el 21 de junio de 2021).

OCDE (2021a). *OECD good practice principles for data ethics in the public sector*. OECD Publishing. <https://www.oecd.org/gov/digital-government/good-practice-principles-for-data-ethics-in-the-public-sector.htm>.

Shin, T. (2020, junio). «Real-life examples of discriminating artificial intelligence». Towards data science. <https://towardsdatascience.com/real-life-examples-of-discriminating-artificial-intelligence-cae395a90070>.

Spielkamp, M. (2017, junio). «Inspecting algorithms for bias». *MIT Technology Review*. <https://www.technologyreview.com/2017/06/12/105804/inspecting-algorithms-for-bias/>.

Stirling, A., Chubb, J., Montana, J., Stilgoe, J. y Wilsdon, J. (2019). *A review of recent evidence on the governance of emerging science and technology*. Commissioned by the Wellcome Trust. https://wellcome.org/sites/default/files/evidence-review-governance-emerging-science-and-technology_0.pdf.

Ubaldi, B., Le Febvre, E. M., Petrucci, E., Marchionni, P., Biancalana, C., Hiltunen, N., Intravaia, D. M. y Yang, C. (2019). «State of the art in the use of emerging technologies in the public sector». *OECD Working Papers on Public Governance*, n.o 31; <https://doi.org/10.1787/932780bc-en>.

WEF (2018a). *Agile governance: reimagining policy-making in the Fourth Industrial Revolution*. Foro Económico Mundial. <https://www.weforum.org/whitepapers/agile-governance-reimagining-policy-making-in-the-fourth-industrial-revolution/>.

WEF (2019a). *AI governance: a holistic approach to implement ethics into AI*. Foro Económico Mundial. <https://www.weforum.org/whitepapers/ai-governance-a-holistic-approach-to-implement-ethics-into-ai/>.

WEF (2020a). *AI procurement in a box. Procurement guidelines*. Foro Económico Mundial. <https://www.weforum.org/reports/ai-procurement-in-a-box/>.

WEF (2020b). *The future of jobs report 2020*. Foro Económico Mundial. <https://www.weforum.org/reports/the-future-of-jobs-report-2020/>.

WEF (2020c). *Reimagining regulation for the age of AI: New Zealand pilot project*. Foro Económico Mundial. <https://www.weforum.org/whitepapers/reimagining-regulation-for-the-age-of-ai-new-zealand-pilot-project/>.

WEF, IMDA y PDPC. (2020). *Companion to the model AI governance framework —Implementation and self-assessment guide for organizations*. Foro Económico Mundial. <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGIsago.pdf>.

Zetzsche, D. A., Buckley, R. P., Arner, D. W. y Barberis, J. N. (2017). «Regulating a revolution: from regulatory sandboxes to smart regulation». *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3018534>.

ck
ages™

iStock
by Getty Images™

iStock
by Getty Images™

iStock
by Getty Images™

ck
ages™

iStock
by Getty Images™





caf.com
@AgendaCAF
transformaciondigital@caf.com