



Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en **Inteligencia Artificial en la Administración Pública Federal**

COMPILADO DE COMENTARIOS DE LA CONSULTA PÚBLICA REFERENTE A LOS PRINCIPIOS Y GUÍA DE ANÁLISIS DE IMPACTO PARA EL DESARROLLO Y USO DE SISTEMAS BASADOS EN INTELIGENCIA ARTIFICIAL EN LA ADMINISTRACIÓN PÚBLICA FEDERAL

Con el propósito de fortalecer el uso responsable y ético de Inteligencia Artificial para el bien común, la Coordinación de la Estrategia Digital Nacional, en colaboración con la Unidad de Gobierno Digital de la Secretaría de la Función Pública y la Subcomisión de Inteligencia Artificial de la Comisión Intersecretarial para el Desarrollo del Gobierno Electrónico (CIDGE), generó propuestas en torno a Principios Generales y la Guía de Análisis de Impacto para el desarrollo y uso de sistemas con elementos de Inteligencia Artificial en la Administración Pública Federal en México.

Estas propuestas se dispusieron a consulta pública del 30 de octubre al 18 de noviembre de 2018, mediante la plataforma de gob.mx/participa. Durante este periodo, se recibieron 51 comentarios: 9 relacionados con terminología, 5 con estrategia, 11 al principio de derechos humanos, 5 sobre bienestar y equidad, 5 sobre transparencia, 14 en torno a responsabilidad y seguridad y 2 sobre metodología. En general, la perspectiva de la ciudadanía es que es necesario impulsar más el uso de esta tecnología. Un gran número de opiniones expusieron la importancia del cuidado en el manejo de los datos personales y la protección de libertades y privacidad. Así mismo se reconoció el énfasis en los mecanismos para promover igualdad e inclusión en los beneficios de inteligencia artificial.

El presente reporte consta de dos partes: la primera resume los comentarios que se recibieron para retroalimentar las propuestas; la segunda expone las versiones 1.0 de los Principios y la Guía, que integran las anotaciones de la ciudadanía y una investigación más extensa.

Es importante resaltar que estas versiones se irán fortaleciendo periódicamente, conforme se analizan e implementan distintas aplicaciones en México, y evoluciona la discusión internacional en este tema.



Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

A. SUMARIO DE COMENTARIOS RECIBIDOS DURANTE LA CONSULTA

A continuación se comparte un resumen de los comentarios más relevantes para cada documento, presentados de acuerdo a las secciones establecidas en dichas propuestas.

a) Propuesta de Principios Generales para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

Comentarios recibidos en la sección de Introducción:

- Incluir una breve explicación o antecedente del término Inteligencia Artificial. *Ej. La IA, es una rama de la ciencia de la computación que comprende el estudio y la creación de sistemas computarizados que manifiestan cierta forma de inteligencia (agentes racionales no vivos). Se trata de la construcción de sistemas capaces de aprender conceptos y tareas por sí solos, de razonar y de llevar a cabo conclusiones lógicas para realizar nuevas tareas.*

Comentarios recibidos sobre la Estrategia en general:

- El gobierno debería ser el pionero en la inclusión de tecnologías emergentes para mejorar los servicios públicos.
- La propuesta de IA debe contemplar:
 - a) Desarrollar un marco de gobernanza adecuado para fomentar el diálogo multisectorial, a través de la creación de una Comisión.
 - b) Mapear los usos y necesidades en la industria e identificar mejores prácticas en gobierno;
 - c) Impulsar el liderazgo internacional de México en la materia.
 - d) Trabajar con expertos y ciudadanos mediante la comisión de Inteligencia Artificial para alcanzar la continuidad de estos esfuerzos durante la siguiente administración.

Comentarios recibidos referentes a los principios:

DERECHOS HUMANOS

- Aclarar qué significa autodeterminación.
- El principio de igualdad es contradictorio con el enfoque de inclusión: si se pretende



Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

maximizar el acceso de grupos vulnerables se generan sesgos en el modelo.

- Es cuestionable si es posible “garantizar” la igualdad.
- Ejemplificar el tipo de datos o modelos que podrían presentar sesgos discriminatorios.
- Los sistemas con base en inteligencia artificial no deberían tomar decisiones, sino funcionar sólo como herramientas para facilitar a las personas la toma de decisiones.

EQUIDAD y BIENESTAR SOCIAL

- Incluir el buen uso de las herramientas de IA para evitar, por ejemplo, persecuciones políticas, espionaje injustificado o similares.

TRANSPARENCIA

- Importante indicar qué se debe hacer en el caso de modelos opacos en los que no es posible explicar el proceso de decisión. ¿Están prohibidos en la APF o cómo deben de abordarse?
- La explicación del funcionamiento del modelo deberá ser coloquial y accesible a todo tipo de público o usuarios.
- Poner a disposición del público el conjunto de datos de entrenamiento, la metodología utilizada para generar el modelo y una explicación coloquial del diseño.
- El código y los datos de entrenamiento deben estar disponibles al público, salvo que contengan datos privados o confidenciales.

RESPONSABILIDAD Y OBLIGACIONES

- Generar análisis sobre la factibilidad y facilidad de falsear o desorientar al modelo, con un ejemplo adversarial.
- Asegurar la actualización de los modelos con la frecuencia necesaria para abordar problemas y fenómenos dinámicos.
- Asegurar el manejo de datos personales de acuerdo a la ley federal de protección de datos personales.
- Incluir en un anexo los instrumentos de orden jurídico aplicables al caso de uso.
- Contemplar aspectos como el acceso a los datos, el consentimiento informado, la calidad de los datos, entre otros.
- Contemplar situaciones de seguridad nacional y el incremento de cibercrímenes.
- Deben prohibirse la generación de perfiles secretos o el uso de una puntuación de propósito general en sus ciudadanos o residentes.



Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

b) Guía para el análisis de impacto en el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

Comentarios recibidos referentes a las Dimensiones a analizar: Uso y gestión de datos:

- Incluir si se aplican métodos de seguridad para la protección de datos sensibles o confidenciales .

Comentarios recibidos referentes al Mapeo del nivel de impacto de las cinco dimensiones:

- Definir una métrica numérica para traducir las respuestas cualitativas a una evaluación cuantitativa de impacto para clasificar en niveles.

Comentarios recibidos referentes a la Guía para el desarrollo y uso responsable de sistemas basados en Inteligencia Artificial en la Administración Pública Federal:

- Elaborar una justificación de cómo se distribuyen los beneficios derivados del sistema (desde económicos hasta quién hace uso y tiene acceso a los datos).
- Identificar el nivel de digitalización que tiene la población objetivo del proyecto que usa el sistema basado en IA.

Principios Generales para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

Versión 1.0

Derivado de los comentarios antes expuestos se presenta la versión 1.0 de los Principios Generales y la Guía de análisis de impacto. Esta versión incorpora la retroalimentación de los ciudadanos.



Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

DERECHOS HUMANOS

- + **Proteger la *voluntad y libertades*** de las personas en la implementación de IA, buscando el respeto por la integridad y el derecho a la autodeterminación de los individuos.
- + **Priorizar la *seguridad, integridad y dignidad humana*** en el diseño y aplicación de herramientas de IA, evaluando y monitoreando los factores que pueden impactar en ello.
- + **Fomentar la *igualdad*** al reducir la posibilidad de sesgos discriminatorios en los datos y modelos utilizados.
- + **Promover la *justicia*** al ofrecer mecanismos operados por personas para revisar, apelar y corregir decisiones o acciones tomadas por sistemas basados en IA.

EQUIDAD y BIENESTAR SOCIAL

- + **Buscar el bienestar social** como la principal meta de la implementación de IA, enfocando los esfuerzos en generar un impacto positivo que contribuya a resolver problemáticas públicas.
- + **Tener un enfoque de inclusión** que fomente el acceso de grupos vulnerables a los beneficios asociados al uso de estos sistemas.
- + **Monitorear y evaluar los impactos** de la implementación de IA para asegurar que cumpla con el propósito para el cual fue diseñada.
- + **Mitigar situaciones de riesgo** e incertidumbre de forma continua y generar estrategias de respuesta a los efectos derivados del uso de IA.

TRANSPARENCIA

- + **Explicar a los usuarios** que interactúan con servicios basados en IA, el proceso de decisión que el sistema toma sobre ellos, los beneficios esperados y los riesgos potenciales asociados a su uso.



Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

- + **Fomentar apertura y confianza** al compartir al público la mayor cantidad de información posible, preservando la confidencialidad requerida, que permita entender el método de entrenamiento y el modelo de toma de decisiones del sistema.
- + **Publicar los resultados** de las evaluaciones del sistema procurando incluir versiones dirigidas a distintos públicos.

RESPONSABILIDAD Y OBLIGACIONES

- + **Determinar responsabilidades y obligaciones** claras de distintos actores sobre el proceso de diseño, desarrollo, implementación y uso de la tecnología.
- + **Establecer mecanismos sólidos de protección de datos** de los usuarios, especialmente atributos sensibles, y notificar sobre posibles eventualidades en el manejo de los mismos.
- + **Asegurar la privacidad** al incorporar, desde el diseño, mecanismos de consentimiento y control personal sobre el uso de los datos.

Guía para análisis de impacto en el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal Versión 1.0

El siguiente instrumento es una herramienta práctica para determinar el nivel de impacto de diferentes sistemas basados en Inteligencia Artificial que sean usados por distintas agencias de gobierno ya sea en sus procesos administrativos o en la entrega de servicios a ciudadanos. Esta guía sugiere:

- a) Dimensiones y elementos a considerar en la valoración del impacto de un sistema.
- b) Las medidas a tomar para asegurar un enfoque de responsabilidad en la implementación de los sistemas.



Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

Metodología

La herramienta, basada principalmente en un ejercicio del Gobierno de Canadá, consiste en un cuestionario que analiza cinco dimensiones: Uso y gestión de datos; Proceso; Nivel de autonomía y funcionalidad del sistema; Alcance en individuos, empresas, sociedad; y Alcance en operaciones del gobierno.

Cada pregunta genera una puntuación de 0 a 2, al cual se añade un multiplicador dependiendo de la cantidad de ámbitos en los que tiene efecto (impacto físico o mental, experiencia de usuario, estándares y regulaciones, objetivos/metas, operación, reputación). Ello produce una puntuación para cada dimensión y se identifica un nivel de impacto global.

Posteriormente, se indican las pautas a seguir para responder a los impactos esperados de manera responsable.

1. Dimensiones a analizar:

[Cada pregunta debe tener una respuesta]

1.1 Uso y gestión de datos

¿Quién recolecta los datos de entrenamiento?

¿Quién recolecta los datos de entrada (inputs)?

¿Se monitorea y mejora la calidad de los datos?

¿Se utilizan datos confidenciales?

¿Se utilizan datos personales como insumo?

¿Los datos incluyen atributos o proxies de raza, género, grupos sociales u otros que se presten a condición de discriminación?

¿Potencialmente existen datos desactualizados o no confiables?

	A	B	C
¿Quién recolecta los datos de entrenamiento?	Internamente	Proveedor directo	Terceros
¿Quién recolecta los datos de entrada (inputs)?	Internamente	Proveedor directo	Terceros
¿Se monitorea y mejora la calidad de los datos?	No	Ocasionalmente	Frecuentemente
¿Se utilizan datos confidenciales?	No	Moderadamente confidenciales	Altamente confidenciales
¿Se utilizan datos personales como insumo?	No		Sí
¿Los datos incluyen atributos o proxies de raza, género, grupos sociales u otros que se presten a condición de discriminación?	No		Sí
¿Potencialmente existen datos desactualizados o no confiables?	No		Sí



Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

- ¿Se utilizan datos no estructurados?
- ¿Hay datos que provienen de otros dispositivos conectados a internet (sensores, drones, etc)?
- ¿Se implementan principios de privacidad desde el diseño (*privacy by design*)?
- ¿Se hace perfilamiento de individuos (*profiling*)?

No	Sí
No	Sí
No	Sí
No	Sí

1.2 Proceso

- ¿Se consultó o probó el sistema con grupos de interés y/o grupos vulnerables?
- ¿El sistema será utilizado por actores distintos a quienes lo desarrollaron?
- ¿Se identifican claramente las responsabilidades y obligaciones de distintos actores en las distintas fases de diseño, desarrollo, mantenimiento y mejora del sistema?
- ¿Se delimita claramente el alcance del sistema?
- ¿Existen indicadores de rendimiento del sistema?
- ¿Existe un programa de capacitación sobre la implementación del sistema?
- ¿Existen mecanismos para brindar retroalimentación sobre el sistema?
- ¿Existen procesos de verificación para probar resultados no intencionales?

A	B	C
Ninguno	Grupos de interés	Grupos de interés y grupos vulnerables
No	No se sabe	Sí
Ninguna	Algunas, pero de manera ambigua	Sí, de manera clara
No		Sí
No		Sí
No		Sí
No		Sí
No		Sí

Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

1.3 Nivel de autonomía y funcionalidad del sistema

¿Qué porcentaje del proceso/servicio es automatizado por el sistema?

¿Con qué alcance permite el sistema la intervención humana?

¿El sistema categoriza individuos u objetos?

¿Qué tipo de aprendizaje usa el modelo?

¿El sistema sugiere cursos de acción para una decisión tomada por un humano o toma completamente una decisión?

¿El sistema toma decisiones que requieren discrecionalidad?

¿El sistema ejecuta acciones automáticamente basadas en sus propias decisiones?

¿El sistema interactúa con otros sistemas de TI?

¿Es posible registrar las decisiones del sistema?

¿Es posible brindar explicaciones sobre las decisiones seleccionadas?

A	B	C
1-33%	34-67%	68-100%
Ninguno, no la permite	Alcance limitado (durante el entrenamiento)	Alto alcance (entrenamiento, ajuste/perfeccionamiento, prueba)
Ninguno	Objetos	Individuos
Supervisado	Semi-supervisado o de por Refuerzo	No supervisado
Sugiere cursos de acción		Toma completamente una decisión
No		Sí
No		Sí
No		Sí
No		Sí
No		Sí

1.4 Alcance en individuos, empresas, sociedad

¿El sistema se usa para predecir potencial criminal de un individuo/ organización?

A	B	C
No		Sí



Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

¿El sistema se usa para definir el acceso a atención médica de una persona?

No	Sí
----	----

¿El sistema se usa para decidir sobre el acceso a fondos para un individuo/comunidad/empresa?

No	Sí
----	----

¿El sistema se utiliza para tomar decisiones específicas sobre grupos vulnerables?

No	Sí
----	----

¿El sistema se utiliza para predecir movimientos y efectos en los mercados?

No	Sí
----	----

¿El sistema se utiliza para emitir permisos, licencias, derechos de propiedad intelectual, entre otros?

No	Sí
----	----

¿Se utilizan datos personales para influir en el individuo (comportamiento, acciones)?

No	Sí
----	----

¿El sistema predice comportamientos a nivel del individuo o predice patrones en poblaciones?

Comportamiento de individuos	Patrones en poblaciones
------------------------------	-------------------------

1.5 Alcance en operaciones del gobierno

¿Podrían las decisiones directas del sistema salvar vidas?

A	B	C
No		Sí

¿El sistema reduce la carga administrativa o laboral?

No		Sí
----	--	----

¿El sistema mejora en general la calidad de las decisiones?

No		Sí
----	--	----

¿El sistema reduce los costos de transacción de un programa existente?

No		Sí
----	--	----

¿El sistema puede realizar tareas que los seres humanos no podrían realizar en un período de tiempo razonable?

No		Sí
----	--	----

Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

¿El sistema mejora la entrega de servicios a los ciudadanos?

No	Sí
----	----

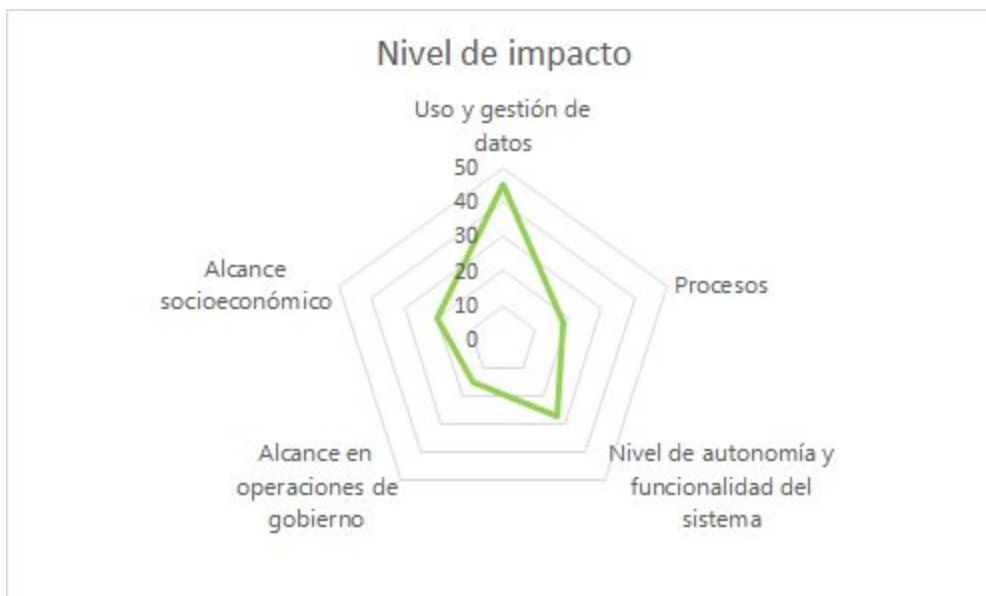
¿El sistema permite desarrollar enfoques innovadores para atender problemas públicos?

No	Sí
----	----

2. Mapear el nivel de impacto de las cinco dimensiones:

Se propone distinguir cuatro niveles de impacto de uso del sistema:

	Bajo	Moderado	Alto	Muy Alto
Uso y gestión de datos	0-12	13-25	26-38	39-52
Procesos	0-7	8-15	16-23	24-32
Nivel de autonomía y funcionalidad del sistema	0-9	10-20	21-31	32-42
Alcance en operaciones de gobierno	0-5	6-11	12-16	17-22
Alcance socioeconómico	0-12	13-25	26-38	39-52





Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

3. Definir el nivel de impacto total del sistema:

Nivel IV: Si dos o más dimensiones, incluyendo alcance socioeconómico, tienen impacto alto o muy alto.

Nivel III: Si el alcance socioeconómico o en gobierno son moderados y dos o más dimensiones tienen impacto alto o muy alto.

Nivel II: Si el alcance socioeconómico o en gobierno son moderados y otra dimensión tiene un impacto alto o muy alto.

Nivel I: Si el alcance socioeconómico y en gobierno son bajos.

4. Atender la Guía para el desarrollo y uso responsable de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

Requisitos	Nivel I	Nivel II	Nivel III	Nivel IV
Antes de operar el sistema:				
Registrar el sistema ante la UGD ¹ , incluyendo descripción clara y completa de su función, objetivos e impacto previsto.	Sí	Sí	Sí	Sí
Firmar los Principios Generales de Desarrollo y Uso de Inteligencia Artificial y Sistemas Autónomos en la Administración Pública en México	Sí	Sí	Sí	Sí
Entregar un reporte a la UGD ¹ con preocupaciones éticas, riesgos y posibles usos no planeados del sistema.	No	No	Sí	Sí
Informar cómo el sistema cumple con los requisitos legales aplicables para operar, incluyendo la legislación sobre protección de datos personales.	Sí	Sí	Sí	Sí
Definir actores y mecanismos de rendición de cuentas particulares al sistema.	No	No	Sí	Sí



Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

Obtener autorización para operar	De la Secretaría correspondiente	De la Secretaría correspondiente	De la Secretaría correspondiente	De la Secretaría correspondiente
Notificar al Consejo interno de Ética el uso del sistema	No	No	Sí	Sí
Asignar recursos para la investigación del impacto e implicaciones de usar el sistema.	No	No	No	Sí
Durante la implementación del sistema:				
Definir y cumplir con estándares mínimos de precisión para implementar el sistema.	No	No	Sí	Sí
Llevar a cabo pruebas de robustez, confiabilidad e integridad del sistema y del modelo.	No	Anual	Semestral	Trimestral
Realizar evaluaciones para mitigar posibles sesgos, incluyendo calidad y relevancia de los datos.	Anual	Anual	Semestral	Semestral
Llevar a cabo auditorías por parte de externos que analicen cómo se llevan a cabo las decisiones.	No	No	Sí	Sí
Capacitación de funcionarios públicos en la función e implementación del sistema para revisar, explicar y supervisar la toma de decisiones automatizada.	Ninguna	Manuales, guías y documentación	Cursos y talleres	Capacitaciones recurrentes y certificadas
Asegurar intervención humana (Human-in-the-loop) en los procesos más críticos o vulnerables del sistema.	No	No	Sí	Sí
Publicar información (variables, metadatos) sobre los datos utilizados en el entrenamiento de un algoritmo y la metodología para el diseño del modelo.	No	No	Sí	Sí
Brindar notificación oportuna al usuario de que es un sistema autónomo con quien interactúa o toma decisiones al brindar el servicio.	No	Sí	Sí	Sí
Compartir al usuario y al público una descripción clara y completa del modelo y su	No	Sí	Sí	Sí



Principios Generales y Guía de análisis de impacto para el desarrollo y uso de sistemas basados en Inteligencia Artificial en la Administración Pública Federal

impacto esperado.				
Reportar los actores que tienen acceso a los datos (entrenamiento y entrada), y evidenciar el consentimiento de los individuos que brindaron datos.	No	Sí	Sí	Sí
Después de usar el sistema:				
Brindar una explicación significativa, clara y oportuna a los usuarios sobre cómo y por qué se tomó la decisión (incluir variables, lógica y técnica).	Por solicitud, en términos generales.	Por solicitud, para el caso específico	Por solicitud, para caso específico y con detalle de variables utilizadas.	Siempre, para caso específico y con detalle de variables utilizadas
Proporcionar mecanismos de apelación a las decisiones automatizadas.	No	Sí	Sí	Sí, con canales directos.
Reportar la actualización de los modelos para abordar problemas y fenómenos dinámicos.	No	Sí, anual.	Sí, anual.	Sí, semestral
Publicar información sobre la efectividad y eficiencia del sistema.	No	Opcional	Sí, anual.	Sí, semestral

¹ UGD: Unidad de Gobierno Digital de la Secretaría de la Función Pública

Referencias:

Principales documentos tomados como referencia para la elaboración de esta propuesta:

- [Future Of Life Institute Published The Asilomar Ai Principles](#)
- [IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems](#)
- [The Engineering & Physical Sciences Research Council \(EPSRC\) Principles of Robotics](#)
- [BS8611:2016 Guide to the ethical design and application of robots and robotic systems](#)
- [The Japanese Society for Artificial Intelligence](#)
- [Universal Declaration of Human Rights](#)
- [European Commission Statement on Artificial Intelligence, Robotics and Autonomous Systems](#)
- [Directrices Universales para la Inteligencia Artificial de la coalición The Global Voice](#)
- [Canada Algorithmic Impact Assessment.](#)
- [Nesta \(2018\). 10 principles for public sector use of algorithmic decision making.](#)
- [Singapore Personal Data Protection Commission \(2018\). Discussion Paper on Artificial Intelligence \(AI\) and Personal Data – Fostering Responsible Development and Adoption of AI](#)