REVISIONES TEMÁTICAS

**Teknokultura.** Revista de Cultura Digital y Movimientos Sociales
ISSNe: 1549-2230

EDICIONES COMPLUTENSE

http://dx.doi.org/10.5209/TEKN.79692

# Especificando la responsabilidad algorítmica

Jorge Francisco Aguirre Sala[1]

Recibido: 6 de enero de 2022 / Aceptado: 27 de marzo de 2022 Open peer reviews

**Resumen.** Especificar la responsabilidad algorítmica tiene por objetivo clasificar las acciones de protección ante los impactos de la Inteligencia Artificial. La descripción de los problemas causados por la Inteligencia Artificial, aunada a la revisión de los modelos y componentes de las evaluaciones, orientan las buenas prácticas y métodos para establecer la huella algorítmica. El análisis de cuatro modelos de evaluación muestra que los mejores modelos son los del riesgo y la responsabilidad legal. Las buenas prácticas de evaluación apuntan a obtener expresiones cuantitativas de aspectos cualitativos, mientras las conclusiones advierten dificultades para construir una fórmula estandarizada. Las métricas de las expresiones cuantitativas deben considerar ponderaciones, según el número de ámbitos afectados, y establecer la gravedad en cuatro niveles de impacto, riesgo o daño. Ello permitiría la reciprocidad de cuatro acciones de protección: prohibir algunos sistemas, asegurar la reparación de daños, promover la mitigación de impactos y establecer la prevención de riesgos.
**Palabras clave:** acciones de protección; evaluación de impacto; inteligencia artificial; modelos de huella algorítmica.

## [en] Specifying algorithmic responsibility

**Abstract.** In seeking to specifying algorithmic responsibility, the aim is to classify protective actions against the impact of artificial intelligence. The article provides a description of the problems caused by artificial intelligence, as well as a review of evaluation models and their components in order to guide best practice and methods in the specification of the algorithmic footprint. The analysis of four evaluation models shows that the best models are those related to risk and legal responsibility. Good evaluation practices endeavor to obtain quantitative expressions of qualitative aspects, while the conclusions warn of difficulties in building standardized formulas. The metrics of quantitative expressions must consider weights, based on the number of areas affected, and establish the severity in four levels of impact, risk or damage. This permits the reciprocity of four protective actions: the prohibition of some systems, ensuring damage repair, promoting impact mitigation, and establishing risk prevention.
**Keywords:** algorithmic footprint models; artificial intelligence; impact evaluation; protective actions.



**Cómo citar:** Aguirre Sala, J. F. (2022). Especificando la responsabilidad algorítmica. *Revista de Cultura Digital y Movimientos Sociales, 19*(2), 265-275. http://dx.doi.org/10.5209/TEKN.79692

[1] Universidad Autónoma de Nuevo León (México)
E-mail: jorgeaguirresala@hotmail.com; https://orcid.org/0000-0002-5805-4082

## 1. Introducción

La Inteligencia Artificial (IA) ha mostrado efectos perniciosos. Sin embargo, no existen legislaciones o certificaciones que unánime y globalmente gradúen con precisión sus peligros y daños. Los debates recientes discuten diversos modelos y metodologías para definir las responsabilidades algorítmicas. Algunas metodologías incluyen las Algorithmic Impact Assessment (Evaluaciones de Impacto Algorítmico, conocidas como AIA, por sus siglas en inglés), que pretenden establecer el grado de la huella algorítmica y los alcances de su carga, similares a las evaluaciones de impacto ambientales que registran la huella ecológica. Las AIA aportarían elementos para evidenciar los impactos, riesgos y daños causados por las decisiones automatizadas de la IA.

Para especificar la huella algorítmica con AIA y establecer niveles de responsabilidad y acciones de protección, se siguen cuatro objetivos. Primero, describir los problemas que causan la necesidad de AIA. Segundo, analizar los modelos existentes de evaluación. En tercer lugar, discernir los elementos que deben contener las AIA. En cuarto lugar, revisar las principales lecciones de las buenas prácticas de evaluación. Estos objetivos se alcanzan con el itinerario consignado en el sumario.

Los resultados de revisar y analizar AIA disponibles muestran heterogeneidad en enfoques de política pública, denominaciones, metodologías y normatividades. La mayoría de estas evaluaciones son realizadas por equipos multidisciplinarios, multisectoriales y dan preferencia a los modelos de responsabilidad algorítmica basados en el riesgo como fundamento de la responsabilidad legal ante los derechos de datos personales, sobre los que tienen otros criterios. Pues «el riesgo se ha convertido en un nuevo límite en el campo de la protección de datos y un indicador clave en decidir si se requieren garantías legales y procesales adicionales» (Macenaite, 2017, p. 507).

Los resultados también advierten dificultades para construir una metodología de AIA con fórmulas estandarizadas. No obstante, las discusiones sugieren transitar hacia expresiones cuantitativas de los aspectos cualitativos. Aquí se proponen acciones de protección correlacionadas a los niveles de daño o riesgo.

### 1.1. Problemas provocados por la IA

La Comisión Especial sobre Inteligencia Artificial en la era digital del Parlamento Europeo (Dalli, 2021) identifica seis conjuntos de inconvenientes. Dichas problemáticas son: 1) riesgos para la seguridad de los ciudadanos, porque no están suficientemente cubiertos por los marcos de protección; 2) riesgo de violaciones a los derechos fundamentales; 3) las autoridades carecen de procedimientos y recursos para garantizar las reglas aplicables; 4) la incertidumbre legal disuade a las empresas de desarrollar y utilizar sistemas de IA; 5) la desconfianza en la IA podría ralentizar su desarrollo y reducir la competitividad global; 6) las reglas fragmentadas crean obstáculos para un mercado único transfronterizo y amenazan la soberanía digital de cualquier nación.

Las AIA no logran atender a todas estas problemáticas. Globalmente es pertinente ahondar en el riesgo de violaciones a los derechos fundamentales y la ausencia de procedimientos legales. El *Libro Blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza* (Unión Europea, Comisión Europea, 2020) reconoce dos dificultades: «Los principales riesgos relacionados con el uso de la IA afectan a la aplicación de las normas diseñadas para proteger los derechos fundamentales […] así como a las cuestiones relativas a la responsabilidad civil» (p. 13). Ambos riesgos impactan a los usuarios finales, quienes terminan por pagar la carga económica del desarrollo algorítmico, debido a la dinámica del mercado establecida entre desarrolladores y usuarios.

La Unión Europea considera que la IA tiene impactos en todos los derechos fundamentales (2021a). Pueden ser positivos, pero también tiene impactos por sesgos, retrasos y errores cuando apoya decisiones humanas o las reemplaza en actividades como la identificación biométrica (con prejuicios raciales), vigilancia, selección de créditos, puestos de trabajo o candidatos a recibir servicios médicos. No es evidente que «[…] estos no son defectos de la tecnología *per se*, sino de la responsabilidad de los humanos que la están diseñando, desarrollando y utilizando» (Unión Europea, Comisión Europea, 2021a, p. 16) debido al autoaprendizaje. A su vez, el asunto de radicar la responsabilidad ha generado múltiples debates sobre quién o qué ha de cargar con la personalidad jurídica garante (Muftic, 2021). De igual manera resulta discutible si los principios de explicabilidad, exactitud, auditabilidad, equidad y declaración de impacto son suficientes para precisar la responsabilidad de los algoritmos (Diakopoulos y Friedler, 2016). A su vez, el asunto de radicar la responsabilidad ha generado múltiples debates sobre quién o qué ha de fincarse la personalidad jurídica garante (Muftic, 2021) y hasta dónde los principios de explicabilidad, exactitud, auditabilidad, equidad y declaración de impacto lograrían fincar responsabilidad a los algoritmos en sí mismos (Diakopoulos and Friedler, 2016), sobre todo cuando los sistemas no son transparentes y los consumidores no pueden detectar los riesgos.

En particular, la dignidad humana es vulnerada cuando al usuario se le hace creer que está interactuando con otra persona y lo hace con un *chatbot*. El algoritmo estereotipa al usuario. Por ejemplo: clasifica a una mujer exenta del riesgo de violencia doméstica; sanciona o libera a un infractor por alta o baja probabilidad de cometer delitos (Hartmann and Wenzelburger, 2021); define a sujetos como inapropiados para una ampliación de crédito o niega una beca.

Los algoritmos también han manipulado las elecciones de un seguro médico o automovilístico según tablas comerciales comparativas, o la adquisición de productos o servicios que implican decisiones bajo diseños subliminales. Son ilustrativos los sesgos algorítmicos en las plataformas de búsqueda de pareja y citas. La IA provoca decisiones al margen de las capacidades cognitivas de los usuarios: por ejemplo, cuando influenció el Brexit o las elecciones presidenciales de los EE.UU. del año 2016. Las recomendaciones personalizadas de productos, como el caso de Amazon o Netflix, segmenta a los usuarios y

provoca fragmentación social, erosionando la cohesión comunitaria y la solidaridad (Yeung, 2019). Peor aun cuando se trata de técnicas no subliminales dirigidas a sectores vulnerables, como lo advierte la Comisión Europea (2021a) «[…]en particular los niños, podrían tener los mismos efectos manipuladores adversos si su condición mental, edad o credulidad se explotan de manera dañina» (p. 17).

Otras afrentas de los algoritmos están en la pérdida de privacidad de datos, sean personales o anónimos como es el caso de la vigilancia masiva y elaboración remota de perfiles biométricos en tiempo real en espacios públicos con fines policiales. En varias ciudades de los EE.UU. ha sido denunciada la práctica policial de detenciones ilegales a personas de etnias de color o rasgos latinos por los resultados algorítmicos de identificación facial, de voz o del modo de caminar (European Union, Agency for Fundamental Rights, 2020). El seguimiento de personas por las vías públicas es una afectación a la privacidad, autonomía, libertad de asociación y reunión, lo cual, también socaba el Estado de derecho, las acciones democráticas y la libertad de expresión. Otros casos que ofenden a los derechos humanos están en la predicción automatizada de las características emocionales. Si los algoritmos pueden reconocer emociones, modificarlas o provocarlas durante las interacciones informativas, entonces incrementan la vulnerabilidad de las personas.

En el sentido anterior, la Comisión Europea (2020, p. 19) consigna «[…] riesgos para la salud mental de los usuarios […] en el caso de la colaboración con robots humanoides». En China destaca el enamoramiento hacia la robot Xiaolce que llevó a psicoterapia a varios usuarios (Entreprenur, 2021). La IA de los robots tiene el poder de funcionar como *influencers* sociales (NeuboxBlog, 2021) instaurando actitudes, valores y decisiones. No es menos sorprendente la erotización de la IA para proporcionar servicios de satisfacción emocional y sexual, como el caso de la robot Harmony (Lean, 2019) que se ostenta como el primer robot sexual personalizable y disponible comercialmente en el mundo. Esto puede verse como perversión, psicopatología o simple entretenimiento tecnológico, pero la sexualidad y el amor entre humanos y robots ha sido inevitable (Levy, 2007). Semejantes fenómenos orillan a los gobiernos, entre los más recientes en 2021, el de China (Del Rio, 2021) y países miembros de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) (2021), a establecer códigos de ética sobre la IA.

La IA puede generar falsos negativos o positivos a nivel individual y colectivo, como lo advierte el anexo 2 del reporte de la European Union, Agency for Fundamental Rights (2020). Un falso positivo existe cuando una persona recibe prestaciones sociales sin tener derecho a ellas. Ese error lacera la calidad de la administración pública y el Estado de derecho. En los falsos negativos, las consecuencias individuales se traducen en pauperización y omisión de recursos que deben otorgarse. En ambos casos el impacto aumenta las desigualdades y la exclusión. Este ejemplo es recurrente dentro de los programas de apoyo a inmigrantes cuando no distinguen entre nacionales y extranjeros por errores de clasificación en apellidos, escolaridad, número de familiares e ingresos.

Existen pocas garantías para que los datos de entrada en un algoritmo sean representativos, completos y depurados de sesgos. Por tanto, los diseñadores no podrían afirmar que los datos de salida estén armonizados con los principios éticos y las legislaciones. Por ejemplo, en el año 2021 un déficit en los sistemas de reconocimiento facial imposibilitaba detectar a las personas en sillas de ruedas.

Los agravios también ocurren en la construcción de perfiles discriminatorios para puntuar procesos de contratación laboral (preferencia de hombres sobre mujeres para trabajos mejor remunerados), de solvencia crediticia (preferencia de nacionales sobre migrantes), servicios de seguros médicos (según perfiles genético-hereditarios) y migrantes (preferencias a quienes tengan familiares en el país receptor). La Unión Europea reconoce que:

> en materia de lucha contra la discriminación, podría ser muy difícil presentar una reclamación, ya que lo más probable es que la persona afectada no sepa que se utiliza un sistema de Inteligencia Artificial e incluso si lo sabe, no es consciente de cómo funciona y cómo se aplican sus resultados en la práctica (Comisión Europea, 2021a, p. 20).

Respecto a las carencias de procedimientos y recursos de las autoridades, las AIA no otorgan poderes *de facto* para garantizar que la IA se ejerza conforme a los derechos humanos y normas armonizadas. Las autoridades también están impedidas en su intervención judicial por los derechos de autor correspondientes a los desarrolladores y propietarios de los algoritmos. Esta condición limita los principios de transparencia, explicabilidad, auditabilidad y justicia. Otras dificultades que tienen las autoridades en atingencia a los derechos de autor son las problemáticas de cooperación entre Estados y las contradicciones o fragmentaciones jurídicas entre legislaciones. Aquello que en un país pude ser obligatorio para las autoridades, en otro puede ser delito, por ejemplo, tomar los datos desde los dispositivos móviles a través de la Corona warn-app para controlar las cadenas de contagio por COVID-19 en China y Corea fue obligatorio, mientras en Alemania está prohibido.

La falta de transparencia, por una parte, hace imposible la defensa de desarrolladores y de las decisiones automatizadas. Por otra parte, dificulta los derechos de audiencia, replica, recurso judicial y un juicio justo e inteligible. Por tanto, como aplicación de los principios éticos deben especificarse las AIA.

## 2. La diversidad de modelos y áreas o dominios de aplicación para establecer la responsabilidad algorítmica

Los modelos para establecer la responsabilidad algorítmica pueden focalizarse en los riesgos desde la formalidad ética, legal o cultural. En la formalidad cultural se atienden las amenazas; en la ética, los diagnósticos y los

daños colectivos; en la legal, se atienden las violaciones a los derechos humanos. Otros modelos surgen desde el interés de los agentes usuarios o consumidores finales o desde la visión de las víctimas, así como desde la asignación de las cargas de responsabilidades. Aiken (2021) propone dos modelos para clasificar los sistemas de IA: por la autonomía y por los riesgos. Sin embargo, para especificar la responsabilidad algorítmica, según Yeung (2019) al menos existen cuatro modelos.

El modelo basado en la intención y culpabilidad se enfoca en la identificación, personalidad jurídica y conducta del agente operador de la IA. Cuestiona si los responsables son los diseñadores y desarrolladores de los algoritmos o son los clientes que indicaron una intención original a estos operadores cuando encomiendan el desarrollo de un sistema. Los dueños y financiadores pueden contratar programadores para que los sistemas ejecuten decisiones reprobables en el ámbito ético, legal o cultural. Los programadores pueden conocer estas consecuencias o no, interponer objeciones de conciencia o simplemente promover el conocimiento informado hacia los dueños, usuarios y consumidores finales con la consabida leyenda «acepto los términos de uso».

El modelo basado en el riesgo y la negligencia indaga daños en perjuicio de usuarios y consumidores durante toda la vida del sistema. Enfatiza la etapa preoperatoria con enfoque preventivo. Propone la carga de obligaciones legales a los sistemas de IA en respuesta a quienes hayan padecido daños por la aplicación y uso. En este modelo surge la discusión sobre el depositario de la personalidad jurídica responsable. En el debate sobre la responsabilidad los argumentos son diversos: hay quienes consideran que los responsables son los dueños y financiadores, otros señalan a diseñadores y desarrolladores y otros más indican que los sistemas de IA en sí mismos (situados en robots androides o no situados) son los responsables debido a su capacidad de autoaprendizaje y autonomía de autoprogramación.

Más allá del debate sobre la personalidad jurídica responsable, el asunto central es la previsión y prevención de los riesgos. No todos los riesgos son previsibles y no todos los daños son causados por la negligencia o a las personalidades jurídicas asignadas de IA. Financiadores, diseñadores y desarrolladores deberían quedar exentos de responsabilidad cuando usuarios y consumidores finales utilizan la IA para fines distintos a los ofrecidos o con acciones allende a las pretendidas originalmente. La previsibilidad de un sistema es importante porque define los propios riesgos y los que están más allá de las limitaciones funcionales y las advertencias originales. Por ende, la prevención y la negligencia pueden implicar a la vez a los diseñadores, los usuarios y los consumidores. De ahí el alto requerimiento de AIA también diferenciado en la vida completa del algoritmo, incluyendo las fases de errores, pruebas de experimentación con respuestas desconocidas y las entradas o salidas de información inusuales. Éstas últimas incluyen los procesos autónomos de autoaprendizaje.

El modelo basado en el riesgo pone en cuestión la responsabilidad de los diferentes actores a lo largo de la vida del sistema. La responsabilidad por defectos de diseño y desarrollo no es equiparable a la causada por nuevas entradas de datos o procesos independientes de autoaprendizaje. Por lo tanto, la AIA debe considerar la fuente de previsibilidad, el riesgo y negligencia, así como el número de ámbitos y sectores donde tienen efectos. Las AIA que utilizan los gobiernos de Canadá (Government of Canada, 2021) y México (Gobierno de México, 2018) ponderan la variedad y el número de sectores de impacto, en una perspectiva contraria a la de Carme Artigas, titular de la Secretaría de Estado de Digitalización e Inteligencia Artificial española, quién afirmó que «los impactos son sectoriales, aunque los algoritmos no lo sean» (citada por Castelló, 2021 y Pascual, 2021).

El modelo de responsabilidad legal se enfoca en los derechos humanos y las libertades fundamentales. La responsabilidad se finca también desde antes de que la norma sea transgredida, por ende, puede prevenirse al tipificar la transgresión y detectarla con las evaluaciones de impacto. Así, este enfoque añade una mejoría al modelo del riesgo: contiene el marco legal que tipifica los daños y previene la vulneración de derechos, protegiendo los datos privados y ordinarios que han sido considerados por los algoritmos. Al respecto, dos especialistas en protección de datos destacan que «el enfoque basado en el riesgo […] se basa en el concepto de que las medidas de privacidad implementadas por los controladores de datos deben ser proporcionales al nivel de riesgos asociados» (Coraggio y Zappaterra, 2018, p. 339).

De manera que «la regulación basada en riesgos implica el desarrollo de marcos de toma de decisiones y procedimientos para priorizar las actividades regulatorias» (Black, 2005, p. 512). En consecuencia, el modelo de riesgo y negligencia y el modelo responsabilidad legal se encuentran mutuamente condicionados. El segundo define y tipifica los daños y riesgos, mientras el primero los advierte o constata. Sin embargo, como la tecnología es más veloz que la jurisprudencia en el desarrollo de nuevos escenarios debe suscribirse lo que indica Macenaite (2017, p. 153) «En un entorno tecnológicamente complejo y que cambia rápidamente, esta postura defensiva basada en el riesgo es cada vez más importante para los reguladores de las nuevas tecnologías, ya que inevitablemente se están quedando atrás en términos de intervención».

La transgresión de los derechos es independiente de cualquier contingencia de los defectos de la IA. Por ejemplo, decisiones automatizadas a partir del autoaprendizaje que resultaron tardías, cuando un sistema pone en manos de humanos la ejecución final con dilación (el control de un vehículo autodirigido, la ingesta de un medicamento prescrito, la expulsión de un inmigrante, el traslado a una prisión de baja seguridad de un preso de alta peligrosidad, etc.). La responsabilidad estricta señala vínculos específicos entre los agentes de IA y las víctimas. Entre los primeros pueden encontrarse financiadores, diseñadores y desarrolladores y, en los segundos, los usuarios finales (instituciones, compañías, comercios y empresas) que interactúan con las personas sujetas a las decisiones algorítmicas. Este último caso puede ejemplificarse con poblaciones vulnerables de

usuarios finales que otorgan becas, prestaciones sociales o empleadores de personal que resultarán riesgosos e inadecuados, aunque el algoritmo de contratación los haya dirigido (quienes contratan a un conductor de trenes, un administrador de empresa, personal de salud en hospitales, celadores en centros penitenciarios o guardias de seguridad privada o pública, etc.).

El modelo de seguro obligatorio se enfoca en la compensación de los afectados por parte de los responsables en vez de concentrarse en la previsión de riesgos o la prevención de daños. Instaurar los seguros obligatorios con cargo a los usuarios de sistemas de IA no siempre podría resultar satisfactorio para los consumidores finales. Este modelo problematiza los costos y las cargas financieras de los mismos. Es decir, la industria de IA alegará que los clientes deben cubrir las pólizas e indemnizaciones, mientras los usuarios exigirán a los productores de los sistemas que costeen los seguros de responsabilidad civil como parte de sus servicios. Como se dijo, el consumidor final al término de la cuenta carga con el costo financiero de los servicios digitalizados.

Existen casos de responsabilidad algorítmica disuelta por la desaparición de los corporativos que construyeron los programas de IA o sus propietarios. Muchas compañías de bienes o servicios otorgaron garantías de por vida, pero la vida de esas compañías o marcas comerciales concluyó antes que dichas garantías fueran reclamadas. El modelo de seguro obligatorio tendría que implementar plazos no perentorios radicados en instituciones, marcas o corporativos, suponiendo que son casi perpetuos para garantizar el cumplimiento. En la práctica también surgió la idea de otorgar personalidad jurídica a los programas de IA situados (el caso de los robots) para cargarles las responsabilidades (Chesterman, 2021; Henz, 2021). El debate sobre la personalidad jurídica de los algoritmos merece un texto aparte.

Las propuestas de responsabilidad legal y seguro obligatorio tendrán que debatir sobre la sustentabilidad económica de los programas para enfrentar los costos de pólizas e indemnizaciones.

De los cuatro modelos anteriores y sus variantes, los de riesgo y la responsabilidad legal son los mejores en cuanto a la capacidad preventiva, la amplitud a lo largo de la vida del sistema, el autoaprendizaje y por evitar costos reparatorios al incluir la prevención y, en el más severo de los casos, la franca prohibición. La complementariedad de ambos modelos es necesaria, ya que, como advirtió Gonçalves (2019, p. 1): «La primera reforma amplia de la legislación de protección de datos personales en la Unión Europea […] ha introducido un enfoque basado en el riesgo […] implementado por los controladores de datos (es decir, los operadores)». En segundo lugar, porque a pesar de que las compañías desarrolladoras de IA podrían incluir programas y sistemas de detección, medición de riesgos y autocorrección, priorizan sus intereses económicos. Aparentan estar preocupadas por salvaguardar los derechos digitales, pero producen sus propias interpretaciones jurídico-tecnológicas sobre qué son y cómo se ejercen los derechos digitales –en particular los atingentes a los datos personales–. La realidad muestra que en la práctica, más allá de las vanas promesas de autorregulación corporativa, las verdaderas soluciones no emergen desde las mismas corporaciones tecnológicas (Vercelli, 2020), como ocurrió en el caso de Cambridge Analytica/Strategic Communications Laboratories, Facebook y el corporativo canadiense AggregateIQ. Estas corporaciones tecnológicas fueron exhibidas por los medios de comunicación en el caso del Brexit y la campaña de Trump en EE.UU. aplicando sus propios criterios jurídicos.

En lo que respecta a las áreas o dominios, Metcalf et al. (2021b) han mostrado la variedad y, por ende, la complejidad, de las diferentes aplicaciones de la IA. Los dominios más regulares son el fiscal, el medioambiental, los derechos humanos, la protección de datos y privacidad. En paralelo deben citarse los dominios a evaluar por algunos instrumentos actualmente parametrizados. El instrumento canadiense considera en lo individual y comunitario a los derechos, la salud, el bienestar e intereses económicos, así como la sostenibilidad del ecosistema y la duración y reversibilidad de los impactos (Government of Canada, 2021). El instrumento mexicano considera los dominios: «derechos humanos, equidad y bienestar social, transparencia, responsabilidad y obligaciones» (Gobierno de México, 2018, p. 1). Se observa que los dominios coincidentes corresponden a los derechos humanos y el bienestar.

La parametrización no sólo corresponde a los dominios, sino también a las dimensiones sobre las cuales habrá que establecer los niveles de impacto. El caso mexicano ilustra las dimensiones de uso y gestión de datos, procesos, nivel de autonomía y funcionalidad del sistema, alcance socioeconómico y de las operaciones del gobierno. Mientras que el instrumento canadiense posee en el cuestionario abierto la categoría de rama, es decir, giro o aspecto a evaluar y aborda las dimensiones sobre el proyecto, el sistema, el algoritmo, las decisiones y los datos. Por su parte, la Unión Europea (2021) incluye otros elementos que privilegian la atención sobre los costos de las evaluaciones y los intereses de viabilidad económica. Como señalan Jiménez y Rendueles (2020, p. 95) la gobernanza algorítmica es: «Más capitalista que digital». Por lo tanto, el estado de la cuestión lleva a preguntar ¿cuáles son los componentes y objetivos que debe poseer toda Evaluación de Impacto Algorítmico?

## 3. Los componentes y objetivos de toda práctica de Evaluación de Impacto Algorítmico

Metcalf et al. (2021b) identifican diez componentes para las evaluaciones de impacto: las fuentes de legitimidad, los actores y el foro, el evento catalizador que detona la necesidad de la evaluación, la temporalidad del sistema, el nivel de acceso público, el método, el conjunto de evaluadores, el impacto mismo y, por supuesto, la determinación de daños y su correspondiente compensación. A dicho listado habría que añadir el nivel de autonomía, la metodología de recopilación de datos y la administración de los inventarios del sistema. El nivel autónomo debe incluirse porque existen sistemas con gran autonomía e impactos despreciables. La recopilación de da-

tos por razones de la invasión a los datos privados o el autoaprendizaje. La administración de los inventarios porque éstos pueden ser temporales o permanentes y, en consecuencia, dejar o no vestigios de responsabilidad.

Según Aiken (2021) establecer los componentes para toda evaluación implica estandarizarla, tal y como se desea con la clasificación de los sistemas. En la opinión de Metcalf et al. (2021b, p. I) esta estandarización sería más adecuada que «[…] la autoevaluación y métricas técnicas estrechas». No obstante, estos autores reconocen que:

> […] analizar e identificar los componentes que necesitan ser agregados o reformados para lograr algoritmos robustos en responsabilidad […] no implica un establecimiento específico de los componentes constitutivos de los AIA […] Estas características también implican que nunca habrá un único proceso de AIA que funcione para cada aplicación o cada dominio (Metcalf et al., 2021b, p. II).

En consecuencia, las AIA sólo pueden incorporar estos componentes de manera limitada. Lo anterior obliga al intento de incluir todas las dimensiones dándoles valor como factor multiplicador de ponderación.

Las fuentes de legitimidad consisten en el conjunto de normas legales con valoración pública que establecen el marco para la rendición de cuentas. Sin estas fuentes no es posible imputar las responsabilidades de los actores vinculados a los algoritmos.

Los actores son los agentes que financian, encomiendan, diseñan, desarrollan y aplican el uso de la IA para la final utilización de los consumidores. Todos ellos son miembros de un foro (implícito o explícito) y comparten responsabilidades. Por consenso deberían tener sus respectivos deslindes.

Un componente relevante es el evento catalizador que detona la necesidad de la AIA. Prever el evento debería ser mandatario por ley para cualquier etapa del sistema. De ahí que los modelos de previsión de riesgos y de responsabilidad sobre los derechos estén mutuamente condicionados. Conocer los riesgos sin la estructura jurídica implicaría un listado caótico de daños y establecer una normatividad sin conocer los riesgos resultará en una estructura vacía.

La temporalidad de las AIA también debería establecerse por ley o por un pacto entre los actores. La AIA debería abarcar desde antes del inicio del desarrollo y la operación hasta los aspectos allende al sistema. Por ejemplo, los diseñadores, desarrolladores y programadores corren riesgos laborales de enajenar su creatividad o convertirse en responsables de daños cuando crean sistemas de autoaprendizaje y aprendizaje profundo capaces de realizar las últimas etapas del desarrollo por sí mismos.

Como es fácil de conjeturar, la amplitud del acceso público a los procesos y documentación de un sistema de IA genera oportunidades para la rendición de cuentas. El acceso debe tener suficiente grado de claridad, explicabilidad, exactitud y auditabilidad. Cuando no hay suficiente claridad y auditabilidad cabe prohibir el uso del sistema, por ejemplo, como los tribunales de Alemania en 2009 y Austria en 2011 prohibieron el del voto electrónico. Los tribunales alegaron la inconstitucionalidad del procedimiento porque el sistema no permitió que cualquier ciudadano no experto en informática pudiera auditarlo. La explicabilidad no siempre es accesible.

Por ende, la consulta pública no debe ser voluntaria, sino obligatoria. La consulta obliga a cumplir las condiciones de acceso transparente y convoca a los actores a precisar los riesgos e impactos. La consulta no es un estudio de opinión o de mercado que oriente el diseño del sistema en términos mercadotécnicos, sino una fuente de investigación para descubrir el peso de la huella algorítmica. Conforme al componente de temporalidad, debe poseer las preguntas pertinentes sobre todas las etapas de los sistemas, no sólo respecto al consumo final.

Especial énfasis merece el método. Las diversas dimensiones, ámbitos y dominios pautan las diferencias entre métodos. No obstante, se buscan características comunes: la determinación del nivel de impacto; la previsión contrafactual; la multidisciplinariedad de datos y recursos; la expresión de procesos y resultados en formatos inteligibles para los actores del foro y la elaboración de dictámenes para la protección legítima de los datos.

El nivel de impacto resulta disímil según las percepciones y legislaciones. Por ejemplo, la evaluación mexicana posee cuatro niveles: bajo, moderado, alto y muy alto. La normativa canadiense distribuye los cuatro niveles en poco o ninguno, moderado, alto y muy alto. Ambos instrumentos agregan dimensiones para obtener el nivel total del sistema. La Unión Europea establece valoraciones más cualitativas y no derivan en niveles.

La estandarización del método es un desafío cuando se desea arribar a métricas en diversos dominios. Como se dijo, deben evitarse métodos disciplinariamente estrechos, es decir, que aborden sólo cuestiones técnicas o sociales, o que los operadores carezcan de experiencia en una amplia gama de impactos. Otro reto es la vinculación del sistema virtual con el mundo real, es decir, evitar que el sistema no alcance aspectos reales de los daños. Si un impacto deviene en daño, este a su vez puede multiplicar más inconvenientes y exceder el conjunto estándar de impactos. Por ejemplo, los dispositivos digitales portátiles en forma de relojes con sensores biométricos (oxímetros, cuenta pasos, cuantificadores de ingesta o quema de calorías, frecuencias cardíacas, etc.), dispositivos de programación deportiva (*fitness,* como los de Fitbit y Nike) y los rastreadores GPS, han controlado las vidas de la mayoría de sus consumidores (De Moya y Pallud, 2020; Ruckenstein y Schüll, 2017).

La multidisciplinariedad obliga que algunos evaluadores sean independientes al grupo de actores en el foro. Las funciones de los evaluadores procuran la transparencia, rendición de cuentas y las decisiones de política pública que normarán a los algoritmos.

El componente central de las AIA es la caracterización del impacto. Las distinciones entre impactos y daños, entre riesgos potenciales y daños, en algunos países permitió la evolución hasta responder a las cuestiones siguientes: ¿qué niveles de impacto han de adoptarse? ¿cuáles son las respuestas legales y las reacciones ins-

titucionales ante cada nivel? ¿quién o qué instancia es responsable de las respuestas y reacciones? ¿cuáles son los medios preventivos, compensatorios o reparatorios aceptables? ¿pueden establecerse escalas de riesgo, daño, compensaciones y reparaciones entre los dominios?

Las AIA se presentan en distintos dominios y sistemas con diferentes grados de avance algorítmico, además de las variadas disposiciones geo-políticas. Por ende, la conmensuración metódica no emerge por sí misma congruente y alineada entre todos los actores debido a los vocabularios, métricas, criterios éticos y cánones legales. Productos de IA diseñados en China, desarrollados en Corea, ensamblados en EE.UU., aplicados en Europa e imitados en Latinoamérica, tienen diferentes impactos y daños. Esa diversidad daña o impide la construcción estandarizada de las AIA. Por ejemplo, la prohibición del servicio Uber en Colombia, por razones de la libre competencia comercial, nunca se hubiera aceptado en el mercado liberal de los EE.UU. Por ende, el método tendrá que ser lo suficientemente abierto, versátil y dúctil para garantizar el cumplimiento de tendencias culturales en diversas instancias, así como la coherencia con distintos sistemas legales de patentes, uso, consumo y disposiciones fiscales. La Unión Europea durante el 2020 y 2021 tuvo que aplicar severas autocorrecciones a la normativa para lograr una versión armonizada (European Union, European Commission, 2021b).

En lo referente a la mitigación de riesgos y reparación de daños, las AIA son rebasadas por la falta de consensos entre los involucrados. A los gobiernos de muchas latitudes (Unión Europea, EE.UU., Australia y varios países de Latinoamérica) les preocupa cargar el costo de las AIA a las pequeñas y medianas empresas desarrolladoras. Los gobiernos se preocupan para que tales situaciones no inhiban el desarrollo de la IA y evitar el rezago en el mediano plazo y la insuperable dependencia de patentes extranjeras en el largo plazo.

Los algoritmos están dentro de la carrera comercial a nivel global, por tanto, el equilibrio entre equidad digital y oportunidades de desarrollo es complejo. Por esta razón, las AIA aplican en su metodología la visión contrafactual. Es decir, ¿qué consecuencias tendría el hecho de prohibir o inhibir un sistema de IA? Sin duda, algunos avances de la IA son imprescindibles, aunque sean cuestionables. Naciones enteras no podrán operar al margen de su uso y caerán en la dependencia de las tecnologías digitales. La pandemia COVID-19 dio una lección de dependencia a muchos dominios cuando, por causa del confinamiento, hubo que recurrir a Google Meet, Zoom, Microsoft Teams y otras plataformas de interconectividad. Para muchos consumidores finales resulta preferible correr los riesgos advertidos por las AIA en vez de quedarse sin IA.

La determinación de daños y la compensación resulta un elemento indispensable. La clasificación por niveles de riesgo y de compensaciones es una primera experiencia. La precisión que ofrecen avanza hacia la categorización con métricas. No obstante, las externalidades compensatorias no resultan simétricas a los riesgos o daños. Por ello, el modelo de seguros obligatorios, además de resultar sospechoso ante el cargo económico prorrateado al consumidor final, no es preferible ante el modelo preventivo de riesgos.

Los dominios también complejizan la adopción de criterios estándar. Así, por ejemplo, el *Libro Blanco* estipulaba:

> Un enfoque basado en el riesgo resulta importante para asegurar que la intervención reguladora sea proporcionada. No obstante, requiere de criterios claros para establecer diferencias entre las distintas aplicaciones de IA, en especial para determinar si entrañan un riesgo elevado o no […] Puede que la legislación de la UE ofrezca categorías de «riesgos» distintas a las que se presentan aquí, en función del sector, como sucede, por ejemplo, en el caso de la seguridad de los productos (Unión Europea, Comisión Europea, 2020, p. 21).

En efecto, los criterios para aplicaciones de seguridad de productos son distintos a los criterios de impacto en los derechos humanos de privacidad de datos o de preferencias de elección.

El Parlamento Europeo se enfoca preferentemente al aspecto económico (Unión Europea, Comisión Europea, 2021c) Aunque posee un objetivo específico de respeto a los derechos fundamentales en su Reglamento y Ley de Inteligencia Artificial. No obstante, su interés primordial es el adecuado funcionamiento del mercado para el desarrollo y uso de la IA. Resulta significativo que decida no definir objetivos operativos de manera detallada.

En contraste, según Golbin (2021), directora responsable de PricewaterhouseCoopers International Limited (PwCIL) en los EE.UU. para la IA, los objetivos específicos que debe perseguir cualquier AIA son**:** 1. Capturar el riesgo de un sistema, 2. Cubrir todos los ciclos de vida del programa, 3. Evaluar el impacto y aumentar la rendición de cuentas a través de un análisis de múltiples partes interesadas y 4. Facilitar las decisiones de continuar o no continuar con el programa.

Metcalf et al. (2021a) definen las AIA como «prácticas emergentes de gobernanza para delinear la rendición de cuentas, hacer visibles los daños causados por los sistemas algorítmicos y garantizar que se tomen medidas prácticas para mejorar esos daños» (p. 735). Sin embargo, existen dos deficiencias en esta concepción: los impactos calculados sólo son reemplazos del daño real y las AIA no son funcionales si carecen de un mecanismo de rendición de cuentas y de reacciones con acciones de protección.

En otras palabras, reconocer impactos no es equivalente a medir daños. Inclusive, detectar algunos impactos puede ocultar ciertos daños. Por ejemplo, impactos benéficos en la detección de cadenas de contagio por el COVID-19 ocultaron los daños al derecho de privacidad o detectar impactos en la selección de personal poliglota para contrataciones laborales oculta discriminación laboral y colonialismo de las lenguas elegidas. De manera que los daños son difíciles de describir y aún más de cuantificar, pues poseen aspectos psicológicos y sociales con grandes complejidades.

Respecto a la rendición de cuentas, éstas no son vinculantes para exigir cambios en los sistemas de IA basados en las AIA, porque no existe un organismo global encargado de la rendición de cuentas algorítmicas. La más reciente recomendación de UNESCO (2021) se declara no vinculante.

Pueden sumarse otras dificultades: los daños suceden dispersos y también agregados, las métricas numéricas inciden en la industria tecnológica pero no establecen la brecha entre la vida real y el poder de los sistemas, no todos los miembros del foro coinciden en las pautas de construcción de criterios de impacto, los encargados de las AIA no tienen accesos adecuados a la información o carecen del expertise técnico de un sistema en particular, las listas de impactos no son definitivas, los monopolios de IA crean sus regulaciones en el marco de sus propias normatividades (taradas por sus intereses) y protegen los aspectos reprobables bajo el derecho de propiedad intelectual. Lo anterior vuelve impracticable la aplicación del principio precautorio de nuevas secuencias integradoras de algoritmos y la estandarización de las AIA.

## 4. Lecciones de las buenas prácticas de Evaluación de Impacto Algorítmico

Los instrumentos para orientar y ejercer la gobernanza sobre los sistemas de IA, y en especial sobre los algoritmos, están lejos de ser uniformes. Alemania posee un sistema de regulación de cinco niveles, Dinamarca tiene un sello de ética de datos, Malta un sistema voluntario de certificación (Unión Europea, Comisión Europea, 2020, p. 12). El Observatorio de Inteligencia Artificial de la Organización para la Cooperación y el Desarrollo Económicos (OECD), en su misión de «realizar una evaluación de impacto y prospectiva tecnológica sobre la IA» recoge, por países y por iniciativas, las propuestas para las políticas de IA (OECD.AI, 2019). En la observación constata muchos tipos de política pública y la heterogeneidad de estrategias, instrumentos, normas, guías, pactos, códigos, enfoques, cánones y un larguísimo etcétera de consideraciones sobre la aplicación y limitaciones de la IA. En el registro al 28 de diciembre de 2021, contabilizó 249 iniciativas de guía y regulación y 479 de gobernanza (OECD.AI, 2021). Otra fuente que constata buenas prácticas es el Observatorio Alemán de la IA en el trabajo y la sociedad (Bundesministerium für Arbeit und Soziales, 2021**)** asociado al OECD.AI.

Del universo concerniente a las AIA, la rendición de cuentas, la transparencia, las guías o los talleres de trabajo –por mencionar los instrumentos más recurrentes en ambos observatorios y los trabajos de la Unión Europea–, destacan como buenas prácticas las de los gobiernos de Canadá, Australia, Japón, Singapur, Reino Unido, los EE.UU. y Nueva Zelanda, (European Union, European Commission, 2021d, p. 33-34). En el sector global gubernamental debe mencionarse el Open Government Partnership que agrupa 78 países y 76 jurisdicciones locales. En el sector no gubernamental destaca la propuesta de Andrade y Kontschieder (2021) que tomó la elaboración de diez compañías mundiales ubicadas en once países, un número significativo de consorcios emergentes dedicados a establecer procedimientos de transparencia y explicabilidad de algoritmos para empresas privadas que son adaptables al sector público. Sobresalen también los trabajos del Ada Lovelace Institute y AI Now Institute and Open Government Partnership (2021).

Aún con el riesgo de marginar algún caso, las buenas prácticas están señaladas por los organismos arriba citados. La evaluación canadiense (Government of Canada, 2021) es reconocida porque ofrece, precisamente, un algoritmo en línea, que obtiene puntuaciones de impacto bruto y de mitigación. Otras prácticas destacables provienen desde 2016 gracias al gobierno francés, al del Estado de California y la *Hoja de ruta hacia un ecosistema de garantía de IA eficaz* ofrecida por el Reino Unido (Gov.U.K., 2021).

Las mejores prácticas aportan las siguientes lecciones: 1) incluir todas las áreas, o dominios de impacto (derechos humanos, salud, vida laboral, etc.). Preferiblemente jerarquizadas en función del alcance demográfico 2); exhortar a todos los involucrados a emitir juicios y sentencias, después de una capacitación y acceso a documentación clave; 3) establecer una metodología accesible, transparente, explicable y que especifique la importancia y escalas de ponderación de diversos elementos de las AIA; 4) aplicar las AIA a lo largo de toda la vida del sistema; 5) diseñar el método de AIA flexible y abierto a la evolución de los algoritmos contemplando el desarrollo de autoaprendizaje y aprendizaje profundo propios de la IA.

## 5. Discusión y conclusiones

En primer lugar, las buenas prácticas sugieren priorizar los modelos de responsabilidad algorítmica basados en el riesgo y la negligencia sobre los que se basan en la intención, la culpabilidad o los que proponen planes de seguro obligatorios. El modelo del riesgo y el de la responsabilidad legal están recíprocamente implicados. Porque advertir riesgos es previsorio y adelantarse a los impactos de negligencias promueve un marco jurídico dinámico. En palabras de Black (2010, p. 303) el riesgo es un «principio organizador significativo de gobierno».

La falta de transparencia muestra la imposible defensa de las decisiones automatizadas y los desarrolladores. Por otra parte, dificulta los derechos de audiencia, réplica, recurso judicial y un juicio justo e inteligible para los afectados. Por ende, no bastan los principios de responsabilidad, auditabilidad y justicia que se reflejan parcialmente en los instrumentos de rendición de cuentas. Expertos externos deben intervenir añadiendo indagatorias multifactoriales y con pesos ponderados. Los debates actuales sugieren transitar hacia expresiones cuantitativas de los aspectos cualitativos y requerir explícitamente una evaluación con métricas de desempeño (Edwards y Veale, 2017).

La variedad y constante desarrollo de los algoritmos muestran severas dificultades para construir una fórmula

estandarizada de AIA (Metcalf, et al., 2021b). No obstante, es deseable la cuantificación de aspectos cualitativos para establecer niveles de impacto, riesgo y daño.

Definir niveles precisos de impacto, riesgo y daño permite establecer en simetría las respectivas acciones. Es decir, se propone que toda AIA contemple al menos cuatro niveles generales para determinar concomitantemente cuatro acciones de protección: la prohibición, la reparación, la mitigación y la prevención. Las AIA serán imprescindibles al reconocer que «necesitamos definir y medir la precisión del modelo, las sanciones y recompensas, los cambios en el desempeño algorítmico debido a la volatilidad ambiental, los niveles de supervisión y sus costos» (Mateos-García, 2017, párr. 37). Es decir, la taxonomía de riesgos y daños evaluados debe implicar acciones de protección proporcionales establecidas desde el modelo de responsabilidad jurídica basada en los derechos digitales.

Pretender la graduación de niveles implica que las métricas sean amplias y con rangos para incluir *ad hoc* a varios sectores y, en lo posible, finalizar con una generalización estandarizada. Por esta razón son importantes las escalas de riesgo más allá de las encuestas y opiniones de los actores involucrados (Kaminski y Malgieri, 2021).

Parece que el debate seguirá abierto respecto al asentamiento de la personalidad jurídica responsable, los costos de las AIA y de las acciones de protección derivadas. Sin embargo, los propietarios o financiadores de los sistemas no tienen la misma condición en el sector público gubernamental que en el sector privado. En todo caso, debería procurarse que los costes no recaigan en los desarrolladores, aplicadores, usuarios y consumidores finales. Andrade y Kontschieder (2021) recomiendan a los legisladores de IA que se centren en el procedimiento en lugar de la prescripción. Ello permite reglamentar requisitos de implementación para las organizaciones que utilizan la IA. También recomiendan incluir justificaciones para las acciones de protección, las cuales deben referir valores éticos y sociales.

Finalmente, por la vertiginosidad de la IA, la necesidad de especificar la responsabilidad algorítmica simultáneamente en varios dominios y, sobre todo, señalar las acciones de protección, es concluyente que el estado actual de las AIA no posibilita una exactitud estandarizada. No obstante, la falta de exactitud no implica ausencia de rigor en la protección de derechos y principios.

## 6. Referencias


Ada Lovelace Institute y AI Now Institute and Open Government Partnership. (2021). *Algorithmic accountability for the public sector.* https://www.opengovpartnership.org/wp-content/uploads/2021/08/algorithmic-accountability-public-sector.pdf

Aiken, C. (2021). *Classifying AI systems.* Georgetown Universtity´s Center for Security and Emerging Technology. https://cset.georgetown.edu/publication/classifying-ai-systems/

Andrade, N. y Kontschieder, V. (2021). *AI Impact Assessment: A policy prototyping experimen*t. Open Loop. https://ssrn.com/abstract=3772500; http://dx.doi.org/10.2139/ssrn.3772500

Black, J. (Autumn 2005). The emergence of risk-based regulation and the new public management in the United Kingdom. *Public Law*, 512-549 http://eprints.lse.ac.uk/id/eprint/15809

Black, J. (2010). The role of risk in regulatory processes. En R. Baldwin, M. Cave, M. Lodge (Eds.), *The Oxford Handbook of Regulation* (pp. 302-348). Oxford University Press.

Bundesministerium für Arbeit und Soziales (2021)**.** *Observatorium Künstliche Intelligenz in Arbeit und Gesellschaft.* https://www.ki-observatorium.de/

Castelló, C. (2021, 23 de julio). España, campo de pruebas europeo para la inteligencia artificial. *Cinco Días. El País, Economía*. https://cincodias.elpais.com/cincodias/2021/07/22/companias/1626964806_533819.html

Chesterman, S. (2021). *We, the robots? Regulating artificial intelligence and the limits of the law*. Cambridge University Press. https://doi.org/10.1017/9781009047081

Coraggio, G. and Zappaterra, G. (2018). The risk-based approach to privacy: Risk or protection for business? *Journal of Data Protection & Privacy*, *1*(4), 339-344. https://doi.org/10.1080/13669877.2018.1517381

Dalli, H. (2021). *Artificial intelligence act*. European Parliament, European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/694212/EPRS_BRI(2021)694212_EN.pdf

Del Río, M. (2021, 5 de octubre). China publica código ético para regular la inteligencia artificial, ¿qué diría Isaac Asimov? *GreenEntrepreneur*. https://www.greenentrepreneur.com/article/389444

De Moya J-F. y Pallud, J. (2020). From panopticon to heautopticon: A new form of surveillance introduced by quantified-self practices. *Information System Journal*, 30, 940–976. https://doi.org/10.1111/isj.12284

Diakopoulos, N. y Friedler, S. (2016). How to hold algorithms accountable. *MIT Technology Review*. https://www.technologyreview.com/2016/11/17/155957/how-to-hold-algorithms-accountable/

Edwards, L. y Veale, M. (2017). Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law & Technology Review 16*(1), 18-84. https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1315&context=dltr https://doi.org/10.2139/ssrn.2972855

Entreprenur (2021, 26 agosto). Usuarios del robot XiaoIce han terminado en terapia por enamorarse de su inteligencia artificial. *Entreprenur.* https://www.greenentrepreneur.com/article/381951

European Union, Agency for Fundamental Rights (2020). *Getting the future right. Artificial intelligence and fundamental rights***.** Publications Office of the European Union. https://doi.org/10.2811/774118

European Union (2021). *Tool #12. Format of the IA report* https://ec.europa.eu/info/sites/default/files/file_import/better-regulation-toolbox-12_en_0.pdf

European Union, European Commission (2021b). *Regulatory scrutiny board opinion. Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts*. EC https://www.eu.dk/samling/20211/kommissionsforslag/kom(2021)0206/forslag/1773317/2379083.pdf

European Union, European Commission (2021d). *Commission staff working document impact assessment. Annexes.* EC SWD/2021/84 final Part 2/2. EUR-Lex – 52021SC0084 – ES – EUR-Lex (europa.eu)

Gobierno de México (2018). *Principios y guía de análisis de impacto para el desarrollo y uso de sistemas basados en inteligencia artificial en la administración pública federal.* Secretaría de la Función Pública. https://www.gob.mx/cms/uploads/attachment/file/415644/Consolidado_Comentarios_Consulta_IA__1_.pdf

Golbin, I. (2021, 28 octubre). *Algorithmic impact assessments: What are they and why do you need them?* PricewaterhouseCoopers US. https://www.pwc.com/us/en/tech-effect/ai-analytics/algorithmic-impact-assessments.html

Gonçalves, M. (2019). The risk-based approach under the new EU data protection regulation: a critical perspective. *Journal of Risk Research 23*(3), 1-14. https://doi.org/10.1080/13669877.2018.1517381

Government of Canada (2021). *Algorithmic impact assessment tool* https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html

Gov.U.K. (2021). *The roadmap to an effective AI assurance ecosystem.* https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem/the-roadmap-to-an-effective-ai-assurance-ecosystem#the-roadmap-to-an-effective-ai-assurance-ecosystem

Hartmann, K. y Wenzelburger, G. (2021). Uncertainty, risk and the use of algorithms in policy decisions: a case study on criminal justice in the USA. *Policy Sciences*, *54*, 269–287. https://doi.org/10.1007/s11077-020-09414-y

Henz, P. (2021). Ethical and legal responsibility for artificial intelligence. *Discovery Artificial Intellegence 1*, 2 https://doi.org/10.1007/s44163-021-00002-4

Jiménez, A. y Rendueles, C. (2020). Capitalismo digital: fragilidad social, explotación y solucionismo tecnológico. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, *17*(2), 95-101. https://dx.doi.org/10.5209/TEKN.70378

Kaminski, M. y Malgieri, G. (2021). Algorithmic impact assessments under the GDPR: producing multi-layered explanations. *International Data Privacy Law, 11*(2), 125-144. https://doi.org/10.1093/idpl/ipaa020

Lean, P. (2019). The extension of legal personhood in artificial intelligence. *Revista de Bioética y Derecho,* 46, 47-66. https://scielo.isciii.es/scielo.php?pid=S1886-58872019000200004&script=sci_abstract&tlng=en

Levy, D. (2007). *Love and sex with robots: The evolution of human-robot relationships*. HarperCollins Publishers.

Macenaite, M. (2017). The "riskification" of European data protection law through a two-fold Shift. *European Journal of Risk Regulation, 8*(3), 506-540. https://doi.org/10.1017/err.2017.40

Mateos-García, J. (2017, 17 de mayo). To err is algorithm: Algorithmic fallibility and economic organization. *Nesta.* https://www.nesta.org.uk/blog/to-err-is-algorithm-algorithmic-fallibility-and-economic-organisation/#_ednref12

Metcalf, J., et al. (2021a). Algorithmic impact assessments and accountability: The co-construction of impacts. *FAccT '21, March 3–10, 2021, Virtual Event, Canada*. https://doi.org/10.1145/3442188.3445935

Metcalf, J., et al. (2021b). *Assembling accountability. Algorithmic impact assessment for the public interest*. Data & Society Research Institute. https://datasociety.net/wp-content/uploads/2021/06/Assembling-Accountability.pdf

Muftic, N. (2021). Liability for artificial intelligence. En Z. Slakoper e I. Tot (Eds.), *Digital Technologies and the Law of Obligations*. (pp. 95-118) Routledge. https://doi.org/10.4324/9781003080596

NeuboxBlog (2021). *Los 5 "robots" más populares que trabajan como influencers*. https://neubox.com/blog/los-5-robots-mas-populares-que-trabajan-como-influencers/

Organización para la Cooperación y el Desarrollo Económicos (OECD.AI). (2019). *OECD AI Policy Observatory.* https://oecd.ai/en/dashboards/policy-initiatives/2019-data-policyInitiatives-24186

Organización para la Cooperación y el Desarrollo Económicos (OECD.AI). (2021). *OECD AI Policy Observatory.* https://oecd.ai/en/dashboards

UNESCO (2021). Proyecto de texto de la recomendación sobre la ética de la inteligencia artificial. En *Informe de la Comisión de Ciencias sociales y Humanas*. https://unesdoc.unesco.org/ark:/48223/pf0000379920_spa

Pascual, M. (2021, 22 de julio). El Gobierno prepara mecanismos para medir el impacto social de los algoritmos. *El País, Tecnología*. https://elpais.com/tecnologia/2021-07-22/el-gobierno-prepara-mecanismos-para-medir-el-impacto-social-de-los-algoritmos.html

Ruckenstein, M. & Schüll, N. (2017). The datafication of health. *Annual Review of Anthropology*, *46*(1), 261–278. https://doi.org/10.1146/annurev-anthro-102116-041244

Unión Europea, Comisión Europea (2020). *Libro Blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza.* Unión Europea. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_es.pdf

Unión Europea, Comisión Europea (2021a). *Commission staff working document impact assessment. Accompanying the Proposal for a Regulation of the European Parliament and of the Council. Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts.* Brussel: EC SWD/2021/84 final Part 1/2. EUR-Lex – 52021SC0084 – ES – EUR-Lex (europa.eu)

Unión Europea, Comisión Europea (2021c). *Propuesta de reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (ley de inteligencia artificial) y se modifican determinados actos legislativos de la unión.* C.E. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52021PC0206

Vercelli, A. (2021). El extractivismo de grandes datos (personales) y las tensiones jurídico-políticas y tecnológicas vinculadas al voto secreto. *THĒMIS-Revista de Derecho 79*, 111-125. https://doi.org/10.18800/themis.202101.006

Yeung, K. (2019). *Responsibility and AI. A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework.* Council of Europe. https://rm.coe.int/responsability-and-ai-en/168097d9c5