Digital Future Society

Gobernanza y algoritmos: riesgos y potencial del uso de la inteligencia artificial en el sector público

Un programa de

GOBIERNO DE ESPAÑA
VICEPRESIDENCIA PRIMERA DE GOBIERNO
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

red.es

MOBILE WORLD CAPITAL BARCELONA

Digital Future Society

Sobre Digital Future Society

Digital Future Society es una iniciativa transnacional sin ánimo de lucro que conecta a responsables políticos, organizaciones cívicas, expertos académicos y empresarios para explorar, experimentar y explicar cómo las tecnologías se pueden diseñar, usar y gobernar, a fin de crear las condiciones adecuadas para una sociedad más inclusiva y equitativa.

Nuestro objetivo es ayudar a los responsables políticos a identificar, comprender y priorizar los desafíos y las oportunidades fundamentales, ahora y en los próximos diez años, en relación con temas clave que incluyen la innovación pública, la confianza digital y el crecimiento equitativo.

Para más información, visite digitalfuturesociety.com

Un programa de

GOBIERNO DE ESPAÑA · VICEPRESIDENCIA PRIMERA DE GOBIERNO · MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL · SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

red.es

MOBILE WORLD CAPITAL BARCELONA

Publicado
Mayo del 2021

Aviso legal
La información y las opiniones expuestas en este informe no reflejan necesariamente la opinión oficial de Mobile World Capital Foundation. La Fundación no garantiza la exactitud de los datos incluidos en este informe. Ni la Fundación ni ninguna persona que actúe en nombre de la Fundación será considerada responsable del uso que pueda darse a la información que contiene.

Nota a la versión en español
Este informe ha sido escrito en inglés y traducido al español. Digital Future Society apoya el uso de conceptos técnicos en español y se esfuerza por encontrar una traducción precisa, siempre que sea posible, sin comprometer por ello el significado original del contenido.

Digital Future Society

Contenidos

1. Tenemos que hablar de la IA

El 2020, un año con más preguntas que respuestas

La inteligencia artificial (IA) fue un tema candente en el 2020, gracias, en parte, a la aparición de productos cada vez mejor desarrollados, capaces de ofrecer servicios útiles ya consolidados, basados en tecnologías de IA. En el año 2020, los sistemas de IA realizaron más tareas rutinarias que nunca, desde planificar trayectos y dar indicaciones paso a paso hasta traducir textos entre diferentes idiomas.

Además, el 2020 se recordará como el año en que la IA pasó a un primer plano debido a numerosas decisiones institucionales de gran impacto. Un ejemplo que creó alarma fue el uso generalizado de sistemas de IA en el sector público, como los sistemas automatizados de toma de decisiones (ADMS, por sus siglas en inglés), para gestionar la concesión de prestaciones sociales, a menudo usando datos de poca calidad y algoritmos de escasa precisión.

Por otro lado, los miedos en torno a la IA están creciendo con la proliferación de los sistemas de reconocimiento facial (SRF) —entre otros usos de la IA— en los espacios públicos, incluso por parte de la policía, algo que resulta poco tranquilizador. La vigilancia innecesaria y las limitaciones y violaciones de los derechos humanos, especialmente en regímenes no democráticos, parecen estar ahora en las frías manos de las máquinas. Esas mismas máquinas "proporcionan a la Administración una capacidad sin precedentes para vigilar a sus ciudadanos e influir en sus decisiones, pero también para interferir en elecciones, dar visibilidad a información falsa y deslegitimar el discurso democrático a escala internacional"[1].

En consecuencia, el año 2020 también estuvo marcado por los debates éticos en torno al uso de sistemas de IA más avanzados para ayudar en la gestión de tareas administrativas, incluidos, entre otros, los sistemas de reconocimiento facial, las predicciones algorítmicas sobre el comportamiento de los ciudadanos e incluso el control de la población. El uso de herramientas de IA por parte de la policía y el ejército también fue objeto de debate, así como los prejuicios discriminatorios aplicados por los sistemas informáticos.

La crisis de la COVID-19 no ha hecho sino exacerbar aún más las amenazas que suponen los sistemas de IA. Las instituciones han tenido que reorientar rápidamente los recursos humanos, crear aplicaciones de rastreo de contactos y adoptar nuevos métodos, totalmente digitales, para realizar labores administrativas y prestar servicios públicos. En este contexto, los riesgos incluyen el caer, aun de forma bienintencionada, en la mala gestión o en la infracción de las normas de protección de datos, al usar registros no anonimizados para

[1] Feldstein 2019

desarrollar herramientas de aprendizaje automático que permitan la detección temprana de determinadas conductas, reales o esperadas. De hecho, ya antes de que la COVID-19 ejerciera más presión sobre las Administraciones, eso es lo que había ocurrido con la mayoría de los usos de la IA dentro del sector público.

La IA se considera a menudo una panacea, pero las complejidades que oculta bajo la superficie representan riesgos que hay que tomar en serio. Es claramente necesario que los responsables políticos comprendan mejor los retos y riesgos que conlleva la implantación de la IA, especialmente en el sector público, a fin de aplicar soluciones que puedan ser realmente beneficiosas para todos.

Qué podemos esperar de este libro blanco

El objetivo de este libro blanco es contribuir a un desarrollo inclusivo de la IA y ayudar a restablecer y reforzar la confianza entre los responsables políticos y los ciudadanos. Esto exige realizar un mayor esfuerzo para comprender mejor los efectos de la IA y para desarrollar algoritmos sobre los que sea posible rendir cuentas y ofrecer explicaciones. Además, se necesitan marcos de evaluación sólidos que puedan valorar no solo la eficiencia sino también los resultados y el impacto socioeconómico de la IA.

En palabras de Stephen Hawking, "es posible que la creación de una IA eficaz sea el mayor hito de la historia de nuestra civilización. ¡O el peor! No lo sabemos"[2].

Este libro blanco recoge, en su cuarta sección, cinco estudios de casos de uso de la IA que han suscitado preocupación por haber dado lugar a una reacción social considerable tras su adopción. Cada uno de ellos fue objeto de un acalorado debate entre políticos, académicos, profesionales y ciudadanos. Todos estos ejemplos proceden de países europeos, aunque también se mencionan ejemplos de otros países a lo largo del documento.

La IA bien podría haber recibido el título de Persona del Año 2020 de la revista *Time*, dada la enorme atención que le han prestado los medios de comunicación, el profundo escrutinio científico al que ha sido sometida y los intensos debates políticos y normativos que se abren en torno a las grandes oportunidades y los inmensos riesgos que plantea. En cualquier caso, tanto en el 2021 como en el futuro, no debemos dejar de hablar de la IA.

En la actualidad, nuestra atención se centra sobre todo en lo que muy acertadamente se denomina *IA estrecha* (o *IA débil*), que es la IA diseñada para realizar una tarea delimitada (por ejemplo, el reconocimiento facial, la búsqueda en Internet o el análisis de ciertos conjuntos de datos). No obstante, la era de la IA solo acaba de empezar.[3] El acelerado ritmo de desarrollo tecnológico plantea la cuestión de qué pasará si muchos investigadores logran el principal objetivo a largo plazo: crear lo que se define como *IA general* (o *IA fuerte*), de manera que los sistemas de IA lleguen a ser mejores que los seres humanos en todas las funciones cognitivas.

[2] Kharpal 2017

[3] Oxford Insights 2020

Por qué Europa es diferente

Este libro blanco analiza sobre todo casos europeos, ya que la Unión Europea (UE), para tratar de limitar los riesgos derivados de la IA, ha adoptado la posición de desarrollar una IA responsable, con una finalidad ética y una tecnología sólida. Son dos componentes esenciales para fomentar la confianza y facilitar su adopción. Tomando como base la comunicación del 2018 IA *para Europa* e inspirándose en las directrices éticas del grupo de expertos de alto nivel sobre la *IA*, el enfoque europeo pretende promover una "IA centrada en las personas", al tiempo que fomenta su adopción y la capacidad tecnológica e industrial en toda la economía y el sector público.[4, 5, 6]

Como se subraya en la Estrategia para el futuro digital de Europa, adoptada en el 2020, la UE espera que la IA mejore significativamente la vida de los ciudadanos y aporte grandes beneficios a la sociedad. Lo hará optimizando la asistencia sanitaria y la agricultura sostenible, haciendo más seguro el transporte y aumentando la competitividad de la industria y la eficiencia de los servicios públicos.[7] A este respecto, el libro blanco de la UE sobre la IA describe un enfoque destinado a crear tanto un "ecosistema de excelencia" como un "ecosistema de confianza", haciendo que los sistemas de IA sean "éticos por su diseño", y propone también un enfoque basado en los riesgos para el régimen regulador.[8]

Según la Comisión Europea, es importante velar por que la regulación sea proporcionada. Se prevé un enfoque escalonado en el que los sistemas de IA de alto riesgo deban obtener una certificación obligatoria antes de acceder al mercado. La clasificación de alto riesgo de la IA depende de lo que esté en juego, teniendo en cuenta si el sector y el uso previsto implican riesgos significativos. Los requisitos reglamentarios propuestos para la IA, que se confirmarán en abril del 2021, profundizarán en estos aspectos y fomentarán el debate internacional.

La Comisión aspira a establecer y motivar un enfoque común para alimentar una modalidad distintiva de la IA que sea éticamente sólida y proteja los derechos de las personas y la sociedad. Se espera que los requisitos reglamentarios de la IA sigan un camino similar al del Reglamento General de Protección de Datos (RGPD), que, pese a la oposición de muchos durante su elaboración, inspiró enfoques similares en todo el mundo.[9]

[4] Comisión Europea 2018a
[5] Comisión Europea 2019
[6] Comisión Europea 2020a
[7] Ibid 2020a
[8] Comisión Europea 2020b
[9] Parlamento Europeo 2016

2. El uso de la IA en la Administración: multitud de incógnitas conocidas

Tradicionalmente, el término *IA* hace referencia a la informática aplicada al desarrollo de máquinas o procedimientos capaces de observar su entorno, aprender y actuar de manera inteligente o proponer decisiones, basándose en los conocimientos y la experiencia adquiridos en ese proceso.[10] Algunos de los usos más típicos son el software de aprendizaje automático o profundo; la automatización robótica de procesos (ARP), como la que se emplea en los asistentes de voz; el reconocimiento de imágenes o voz y la traducción de textos, y los sistemas automatizados de toma de decisiones (ADMS, por sus siglas en inglés). También es posible integrar la IA en dispositivos de hardware, como robots avanzados, sistemas autónomos, y sistemas y dispositivos del Internet de las cosas (IoT).

El uso de sistemas y herramientas basados en IA para respaldar la toma de decisiones, la implementación y las interacciones ya está presente en la labor de la mayoría de las Administraciones de todo el mundo. Y es que ofrece un claro potencial de reducir el coste de las funciones básicas de las instituciones, como la aplicación de los mandatos normativos y la adjudicación de prestaciones y privilegios.[11] Sin embargo, muchos casos de uso incluyen también otras tareas de gobernanza sumamente importantes, como el análisis normativo, la elaboración de normas, la gestión del personal interno, la participación de los ciudadanos y la prestación de servicios.

En la mayoría de los casos, los sistemas de IA sirven para mejorar la eficiencia de las Administraciones mediante el análisis automático de enormes volúmenes de datos. Se da por supuesto que ofrecen una perspectiva más completa y precisa que los análisis realizados por personas. Sin embargo, esto no es necesariamente así, ya que los resultados de los análisis de datos informatizados dependen de la calidad de los datos disponibles y de la precisión de los algoritmos empleados.

Pero, además de los problemas y retos que conocemos, la "información conocida", y dejando a un lado las numerosas "incógnitas desconocidas", las características inherentes a la IA y las propiedades de aprendizaje que presentan ponen de relieve la existencia de numerosas "incógnitas conocidas", es decir, los retos y problemas que sí conocemos, pero no sabemos cómo resolver. Esto significa que es urgente abordar las limitaciones actuales de la IA, así como las consecuencias negativas y los efectos secundarios que puede tener para los ciudadanos el uso inadecuado de sistemas de IA.

[10] Craglia et al. 2018

[11] Engstrom et al. 2020

En principio, la IA posee el potencial de mejorar la vida de las personas procesando ingentes cantidades de datos, ayudando a los funcionarios en los procesos de toma de decisiones y proporcionando aplicaciones a medida y servicios personalizados.[12] Sin embargo, la IA también puede aumentar el isomorfismo institucional y consolidar sistemas y estructuras de poder disfuncionales. Si a un sistema disfuncional o un conjunto de datos sesgados se le añade una capa de IA o aprendizaje automático, eso solo servirá para empeorar los problemas preexistentes. Por otro lado, el sector público está expuesto a un escrutinio más profundo, debido al papel y las funciones de las instituciones y al riesgo de intensificar las asimetrías de poder entre los responsables políticos y entre unos ciudadanos y otros.[13]

Como muestran algunos de los ejemplos que se presentan más adelante, los procesos de digitalización suelen afectar a áreas que tratan con ciudadanos en situaciones muy vulnerables, lo que refuerza la necesidad de comprender los riesgos que conlleva el despliegue de la IA en el sector público. Además, hay otras amenazas importantes inherentes a las propiedades de la IA, como las consecuencias de que una máquina deniegue un derecho a través de un sistema de IA, la falta de competencias digitales de los funcionarios o la manera en que funcionan realmente estos sistemas y las implicaciones que tienen para los usuarios.

Los trabajos anteriores de Digital Future Society ponen de manifiesto algunos de los principales retos que supone la introducción de sistemas de IA en el sector público. Dichos retos incluyen la "discriminación por defecto" y el sesgo inherente que puede generar la falta de calidad de los conjuntos de datos sobre la vida de los grupos vulnerables y las personas desfavorecidas; la obstinada opacidad que rodea al creciente uso de sistemas para el llamado "estado de bienestar digital", y el profundo impacto que pueden tener esos sistemas en la relación entre los sistemas democráticos y la "gobernanza algorítmica", teniendo en cuenta el poder de vigilancia que pueden ofrecer estas tecnologías a las instituciones del sector público.[14, 15, 16, 17]

Discriminación por defecto

La IA ofrece a las instituciones diversas oportunidades, pero también plantea numerosos retos. Por ejemplo, aunque puede ayudar a agilizar operaciones y procesos administrativos, también puede resultar inexacta e interferir en la interoperatividad entre departamentos de la Administración. La IA puede mejorar la recopilación de información y ayudar a extraer conclusiones útiles, aplicando análisis predictivos avanzados, pero también tiende a ser invasiva y a menudo puede arraigar aún más los prejuicios sociales e institucionales.

Algunos ejemplos controvertidos son los casos de vigilancia policial predictiva, en los que las fuerzas del orden utilizan tecnologías de IA para tomar decisiones sobre la aplicación de la prisión preventiva o sobre sentencias judiciales, o bien para identificar las zonas en las que es más probable que se produzcan delitos.[18, 19]

[12] Algorithm Watch y Bertelsmann Stiftung 2020

[13] Kuziemski y Misuraca 2020

[14] Digital Future Society 2020a

[15] Alston 2019

[16] Algorithm Watch y Bertelsmann Stiftung 2020

[17] Digital Future Society 2020b

[18] Big Brother Watch 2020

[19] Dencik et al. 2019

En **Estados Unidos**, la **herramienta COMPAS** (siglas en inglés de "creación de perfiles de delincuentes para la aplicación de sanciones sustitutivas en gestión penitenciaria") ofrece el que probablemente sea el caso más notorio de prejuicios en la IA.[20] Desde el 2010, la IA se emplea a menudo de manera similar en varias jurisdicciones de Estados Unidos para predecir la probabilidad de reincidencia de un delincuente. En el 2016, un estudio de ProPublica denunció que, "según las predicciones del sistema, los acusados negros presentan un riesgo de reincidencia mayor que el índice real, y sucede lo contrario con los acusados blancos"[21].

Un estudio posterior demostró que ProPublica cometió un error importante al procesar los datos, que afectó en parte a los valores predictivos positivos y negativos. A pesar de ello, la organización sin ánimo de lucro afirmó que "esto tuvo poco impacto en algunos de los demás parámetros clave de las estadísticas, que son menos susceptibles a los cambios en la proporción relativa de reincidentes, como las tasas de falsos positivos y falsos negativos, y la precisión en general"[22].

Los sistemas de IA que tratan de identificar las zonas de mayor delincuencia se han encontrado con los mismos problemas. Estos sistemas influyen en los agentes de policía que patrullan en las zonas identificadas y los hacen más propensos a detener o arrestar a personas, guiándose por las expectativas suscitadas por el análisis y la predicción del sistema, más que por las circunstancias reales.[23] Cada vez hay más pruebas que sugieren que, al entrenar los modelos de aprendizaje automático con datos policiales sesgados, los prejuicios humanos se refuerzan y se consolidan en los sistemas de IA.[24]

Entre falsos positivos y falsos negativos

Los algoritmos de predicción pueden cometer errores. En el contexto de las tecnologías de reconocimiento facial, por ejemplo, hay dos errores posibles: los falsos positivos, en los que el algoritmo determina una coincidencia positiva entre dos imágenes faciales cuando en realidad no coinciden, y los falsos negativos, en los que el algoritmo concluye que no hay ninguna coincidencia cuando sí la hay.[25]

Un caso que suscitó gran preocupación fue el uso de un **sistema de reconocimiento facial (SRF)** en la ciudad de **Buenos Aires**. Tras una resolución de abril del 2019, el Ministerio de Justicia y Seguridad de la Ciudad Autónoma de Buenos Aires empezó a utilizar un SRF en directo para identificar a niños acusados de cometer delitos.[26] Human Rights Watch (HRW) criticó el sistema y pidió al Gobierno de la ciudad y al nacional que dejaran de utilizarlo para identificar a sospechosos, en particular a menores, señalando que el sistema, a menudo, los identifica erróneamente.

[20] Douglas Heaven 2020

[21] COMPAS es un programa informático usado como herramienta de apoyo para predecir el riesgo de reincidencia, es decir, la probabilidad de que un acusado/da vuelva a delinquir.

[22] Barenstein 2019

[23] Babuta y Oswald 2019

[24] Richardson et al. 2019

[25] Agencia de los Derechos Fundamentales de la Unión Europea 2019

[26] Bronstein 2020

Dicha ONG argumentó que esos errores de identificación podían limitar injustamente las oportunidades laborales y educativas de los niños acusados de robo y otros delitos. Además, se publicaba en Internet la información personal de los niños acusados de haber cometido un delito, lo que va en contra del derecho internacional.[27]

El debate global en torno a los SRF es importante, ya que este uso invasivo y potencialmente perjudicial de las herramientas de vigilancia masiva se está implementando cada vez más en Latinoamérica. Los Gobiernos de Brasil y Uruguay, por ejemplo, han impulsado un marco legal para gestionar el uso de dichos sistemas.

El estado de bienestar digital y sus efectos de *cajanegrización*

También es controvertido el uso de tecnologías de IA para ayudar a las instituciones a detectar anomalías en grandes conjuntos de datos. Por ejemplo, estas tecnologías utilizan los datos para descubrir automáticamente conductas fraudulentas relacionadas con prestaciones de servicios de la Administración, como subvenciones, prestaciones sociales o impuestos (como veremos más adelante), o para identificar a niños y familias considerados vulnerables y en riesgo de sufrir abusos.

Un caso muy discutido fue el del **EHPS** (siglas en inglés de "sistema de creación de perfiles para la ayuda temprana") desplegado por el consejo del **distrito londinense de Hackney**.[28] El sistema se concibió con la idea de ayudar a los consejos municipales a ahorrar alrededor de un millón de libras esterlinas al año, al facilitar la intervención temprana en casos concretos, pero acabó siendo muy criticado por el tipo de datos que recogía y la opacidad de su método de evaluación de riesgos. A la ciudadanía le preocupaba también el hecho de que, aparentemente, el sistema solo se había implementado para cumplir las medidas de austeridad del Gobierno del Reino Unido, ya que según lo anunciado optimizaría los costes del programa Troubled Families ('Familias con problemas').[29]

Finalmente, el consejo municipal de Hackney detuvo el proyecto, al considerar que no se habían obtenido los beneficios esperados.[30] En gran medida, esta historia es ilustrativa de cómo el hecho de centrarse principalmente en la eficiencia y la rentabilidad ha afectado a la digitalización del estado de bienestar en la mayoría de los países. Se piensa poco en el diseño de los sistemas basados en IA, en cómo hacer frente a la falta de transparencia o en el sesgo de los datos que se utilizan para entrenar los algoritmos.[31, 32]

[27] Alston 2020

[28] Dencik et al. 2018

[29] GOV.UK, (s.f.)

[30] Dencik et al. 2019

[31] Douglas Heaven 2020

[32] Digital Future Society 2019

Los sistemas de IA que se utilizan para ayudar a gestionar las solicitudes de asistencia social o calcular las prestaciones sanitarias también han mostrado indicios similares de sesgos sociales y discriminación racial o étnica.[33] El problema radica en que es difícil discernir de dónde pueden provenir los sesgos, ya que los algoritmos son a menudo de propiedad privada y, por tanto, están cerrados al escrutinio. Esto supone un obstáculo adicional, vinculado a la limitada capacidad de los organismos del sector público y la falta de aptitud de los funcionarios para tratar con sistemas tan complejos. Con frecuencia, las personas que trabajan con estos sistemas acaban confiando en las decisiones sugeridas por el ordenador, sin ser capaces de cuestionar o comprender plenamente la lógica que los sustenta.

En la práctica, hay diversas incógnitas conocidas que están surgiendo como cuestiones fundamentales que deben abordar los responsables políticos. Dichas incógnitas demuestran que existe una necesidad urgente de asegurar que los procesos de toma de decisiones y los sistemas institucionales se centren en las personas y se pueda rendir cuentas de ellos; que garanticen la transparencia y la calidad de la gestión y prestación de los servicios públicos, y, en última instancia, que generen bienestar para todos.

[33] Eubanks 2018

3. Gobernanza de, con y por la IA

Ha quedado claro que los responsables políticos se enfrentan a un difícil dilema: la obligación de proteger a los ciudadanos de posibles daños debidos al uso de algoritmos está en conflicto con la tentación de aumentar la eficiencia y mejorar la calidad de los servicios digitales.[34] Deben afrontar dos retos: por un lado, la gobernanza de la IA, es decir, regir los algoritmos y los procesos automatizados relacionados, y por otro, la gobernanza con la IA y por parte de la IA, utilizando los algoritmos y los métodos y sistemas informáticos para mejorar los servicios públicos.

Gobernanza de la IA

Como ocurre con cualquier innovación tecnológica, la introducción de la IA en el sector público no es un proceso sencillo. No deben anularse los mecanismos e instituciones de gobernanza existentes. Deben abordarse las tradicionales barreras tecnológicas, legales y reglamentarias, así como las cuestiones éticas y sociales. También hay otros factores relacionados con la IA, como las inversiones a largo plazo, las habilidades y capacidades, el valor percibido, la sostenibilidad y las dificultades de desarrollar operaciones y servicios digitales básicos en la Administración. Por ello, el tipo de gobernanza "de la IA" que se adopte es fundamental, y no es tan fácil determinarlo de antemano.

La combinación de inmensas cantidades de datos con potentes algoritmos de aprendizaje automático es lo que está impulsando el desarrollo de la IA. Por lo tanto, no se puede hablar de la gobernanza de la IA sin examinar primero los regímenes y prácticas de regulación de datos existentes. Sería lógico establecer la gobernanza de la IA como una extensión de la regulación en materia de protección de datos y competencia. Sin embargo, y por desgracia, la actitud actual en torno a la IA se basa en la narrativa del excepcionalismo, de modo que la IA se considera un fenómeno nuevo, ajeno a las políticas y las leyes existentes.

Esto significa que las Administraciones deben, en primer lugar, comprender mejor las implicaciones normativas y los mecanismos de gobernanza que están cambiando el funcionamiento de las organizaciones y empresas públicas y privadas, así como sus efectos sobre los derechos de los ciudadanos. Solo entonces estarán en condiciones de explorar posibles usos innovadores de las tecnologías que consideren necesarias. Los casos SyRi y Gladsaxe, presentados en la cuarta sección de este libro blanco, ilustran esta cuestión en mayor profundidad.

[34] Kuziemski y Misuraca 2020

Gobernanza con la IA

Otro aspecto importante —aunque a menudo ignorado— que se debe analizar y evaluar es el valor y la utilidad que puede ofrecer la IA a las instituciones cuando rediseñan procesos administrativos internos para mejorar la calidad y el impacto de los servicios públicos.[35]

Gobernar "con la IA" significa que las personas deben ocupar su posición clásica, es decir, utilizar y controlar una tecnología que refuerce su capacidad mediante un proceso que requiera supervisión humana. No obstante, lo más importante es comprender mejor los posibles beneficios y riesgos asociados al uso de la IA en el sector público. Entre ellos se encuentran la protección de los derechos humanos y la implementación ética de la IA, especialmente en cuanto a aspectos políticos delicados y temas de interés social que tengan implicaciones directas y graves en la relación de confianza entre las Administraciones y los ciudadanos.

En **Polonia**, por ejemplo, tanto la población como los funcionarios criticaron un **sistema algorítmico de creación de perfiles** introducido como parte de una reforma de los servicios públicos de empleo, PUP (Powiatowe Urzędy Pracy). El sistema dividía a los ciudadanos desempleados en tres categorías, cada una de las cuales les asignaba un nivel determinado de carga asistencial y de recursos. Fue criticado por su gran opacidad, ya que los ciudadanos desconocían la puntuación que se les otorgaba y cómo se había determinado dicha puntuación.

Además, la idea era que el sistema de perfiles sirviese únicamente como herramienta de asesoramiento, y el personal debía decidir si la categoría asignada era correcta. Sin embargo, en la práctica, el personal interno cuestionó menos del uno por ciento de las decisiones del algoritmo por falta de tiempo, por miedo a las repercusiones dentro de la Administración y por la presunta objetividad del sistema de IA.

Finalmente, se determinó que el sistema era inconstitucional y el Gobierno lo eliminó tras varias quejas formales por sus efectos discriminatorios. Este caso muestra de manera clara que, aunque la participación de personas en el proceso (*human in the loop* o HITL) puede ofrecer una solución, esas personas deben estar capacitadas para cuestionar las decisiones de la IA, especialmente si los sistemas se han introducido para ayudar a ahorrar costes y mejorar la eficiencia. En la cuarta sección de este libro blanco se analiza otro caso similar que tuvo lugar en Austria.

[35] Como se explica en la revisión bibliográfica de Desouza et al. (2020), la investigación sobre la adopción de la IA se centra, casi exclusivamente, en el desarrollo y la aplicabilidad de la IA en el sector privado. Muy pocos de los artículos publicados entre los años 2000 y 2019 (59 de 1.438) se ocupan de la IA en el sector público.

Gobernanza por la IA

En cualquier caso, el verdadero potencial del uso de la IA en el sector público —y los riesgos que conlleva— reside en la gobernanza "por parte de la IA", que implica que las personas responsables de tomar decisiones se rindan a las "capacidades sobrehumanas" de la IA.

Aunque las aplicaciones de este tipo de sistemas de IA están aún en su fase inicial, sobre todo en la Administración, ya estamos asistiendo al rápido desarrollo de sistemas inteligentes/autónomos que no se limitan a ejecutar instrucciones o tareas predefinidas. Estas aplicaciones de IA, más sofisticadas, no dependen de la intervención humana y son capaces de aprender y adaptarse por su cuenta. Pueden usarse como una herramienta colaborativa para identificar problemas, encontrar nuevas soluciones y ejecutarlas más rápidamente y de formas innovadoras. Pero, si se utilizan con malas intenciones, también pueden causar daños o influir en las capacidades cognitivas de los seres humanos, lo que a su vez afectaría profundamente al mundo en que vivimos, tanto a los espacios personales como a los entornos sociales.

Esta evolución agrava aún más las tensiones existentes debido a las desiguales relaciones entre las personas registradas y las que analizan esos datos (ya sean humanos "aumentados" con funciones asistidas por ordenador, o un ordenador sin intervención humana). Esto se ha denominado *riesgo de algocracia* (palabra que deriva de *algoritmo*).[36]

La IA puede aportar mejores resultados para todos, pero antes de embarcarse en una transformación potencialmente radical de la forma de diseñar y prestar políticas y servicios, hay que tener en cuenta los posibles riesgos, consecuencias y efectos secundarios no deseados. En particular, los desafíos relacionados con la rendición de cuentas y la confianza, pero también con la responsabilidad. ¿A quién se responsabilizará cuando un sistema de IA provoque daños a causa de accidentes o errores?

Como ilustran los casos de la siguiente sección, es necesario abordar la cuestión fundamental de cómo las Administraciones diseñan y gestionan los sistemas de IA (¿o son los sistemas de IA los que gestionan a las Administraciones?). También el papel de los proveedores del sector privado, que a menudo controlan los datos y los procesos de los sistemas automatizados de toma de decisiones.

A este respecto, hasta ahora se han publicado pocos estudios empíricos sobre el uso de modelos algorítmicos en la formulación de políticas, lo cual limita la comprensión académica de dicho uso y sus efectos. Por ello, es necesario un esfuerzo específico, en el ámbito político y en el de la investigación, para que el uso de la IA en el sector público reciba más atención.[37]

[36] Danaher 2016

[37] Kolkman 2020

La IA se concibe como un importante motor de cambio para los sistemas de gobernanza, ya que puede permitir un giro paradigmático en las relaciones de poder entre las partes implicadas. Sin embargo, este cambio suele estar impulsado por enfoques tecnodeterministas. Tal y como advierte Evgeny Morozov, "en lugar de corregir las estructuras de asistencia social o las verdaderas causas de las crisis, los solucionistas despliegan la tecnología para evitar la política, y exploran cada vez más formas de influir en nuestro comportamiento para hacer frente a los problemas".[38]

Las implicaciones legales y éticas del uso de la IA (ya sea con ella o por parte de ella) son de vital importancia para garantizar la legitimidad y la fiabilidad de las instituciones y la prestación de unos servicios públicos justos e inclusivos. Al mismo tiempo, el sector público desempeña un papel fundamental a la hora de definir los mecanismos reguladores y las soluciones técnicas para el desarrollo de sistemas basados en la IA en toda la sociedad.[39]

[38] Morozov 2020

[39] Misuraca y van Noordt 2020

4. Qué podemos aprender de los ejemplos europeos de uso de la IA en la Administración

En muchos países, las Administraciones están experimentando con la IA para mejorar la formulación de políticas y la prestación de servicios. Las repercusiones que esto tiene se observan en diversos aspectos del sector público, y a menudo se da por supuesto que dichos efectos son positivos. Sin embargo, existen numerosos ejemplos de uso indebido y de las consecuencias negativas y los daños que puede causar el uso de IA en la Administración. Entre ellos hay varios casos ampliamente difundidos que han suscitado un enorme interés social, con los consiguientes debates políticos.

Tal como se preveía, son muchos los retos relacionados con el uso eficaz de la IA en el sector público, lo cual es un obstáculo para aplicarla de forma generalizada. Una cosa es que un sistema de IA se equivoque prediciendo la palabra que queremos escribir en el móvil: puede molestarnos un poco, pero nada más. Sin embargo, otra muy distinta es que no cumpla con su función si se ha diseñado para ayudar a tomar decisiones sobre la salud de una persona o las prestaciones sociales que debe recibir, por ejemplo.

Desde esta perspectiva, los siguientes casos, observados en Europa, ilustran algunos de los principales riesgos que supone el uso de la IA en áreas cruciales de los servicios públicos y la formulación de políticas. Estos estudios de casos se centran en los países europeos porque, como se ha mencionado anteriormente, la posición de la UE es desarrollar una IA responsable con una finalidad ética y una tecnología sólida.

De la detección de fraudes a la dimisión del Gobierno

Systeem Risico Indicatie (SyRi) es un sistema de IA empleado por varios **municipios holandeses** y el Gobierno nacional "para prevenir y combatir el fraude en los ámbitos de la seguridad social y los regímenes relacionados con la renta, las contribuciones fiscales y de seguridad social, y las leyes laborales"[40]. Fue noticia cuando el tribunal del distrito de La Haya sentenció, a principios del 2020, que no se ajustaba al artículo 8 del Convenio Europeo de Derechos Humanos (CEDH), que estipula que toda persona tiene derecho al respeto de su vida privada, por lo que se deben sopesar los beneficios de las nuevas tecnologías.

[40] Rechtbank Den Haag 2020

SyRi enlazaba y analizaba datos de varios organismos públicos, y generaba un informe de riesgos para ayudar a atajar el mal uso de los fondos y detectar los fraudes. Aunque tenía un fundamento jurídico y se basaba en información clara sobre qué datos podía captar, almacenar o compartir entre los distintos departamentos, el uso del sistema fue muy controvertido, ya que se dirigía y examinaba sobre todo a los ciudadanos pobres y vulnerables, que supuestamente eran más propensos a cometer fraudes.

Una coalición de diversas organizaciones de la sociedad civil y un gran sindicato se quejaron de que el sistema era injusto, ya que no examinaba a todos los ciudadanos por igual y solo se utilizaba en los barrios desfavorecidos: "Si solo buscas en ciertos lugares, solo encontrarás algo en esos lugares"[41].

Esto puso de manifiesto el hecho de que podían establecerse vínculos involuntarios basados en sesgos, como los relacionados con el estatus socioeconómico más bajo o el origen migratorio, especialmente si se tiene en cuenta que los métodos de modelado de datos no estaban abiertos al escrutinio.

Los problemas que puso de relieve el caso SyRi se amplificaron con el reciente escándalo de las autoridades fiscales holandesas. El sistema secreto **Fraude Signalering Voorziening (FSV)** dio lugar a análisis de riesgo incorrectos que llevaron a considerar erróneamente que numerosas personas habían cometido fraudes. Tras varias quejas, las investigaciones demostraron que el sistema utilizaba datos restringidos para detectar señales de posibles fraudes, incluyendo entradas registradas en el FSV que no tenían distinciones de significado y gravedad, lo que llevó a incluir información incompleta, incorrecta y desactualizada.[42, 43]

Este mosaico de prácticas no solo incumplía el RGPD, sino que también daba lugar a prácticas de trabajo poco claras e incorrectas por parte de los funcionarios que manejaban los datos del FSV.[44] Dadas las malas prácticas resultantes, el informe de una comisión parlamentaria de investigación concluyó que se habían "infringido los principios fundamentales del Estado de derecho" al reclamar la devolución de ayudas para el cuidado de los hijos a padres identificados como defraudadores por errores menores, como la falta de una firma en la documentación.[45] Las familias, a menudo de grupos minoritarios y de origen inmigrante, se vieron obligadas a devolver decenas de miles de euros sin posibilidad de subsanación, lo que sumió a muchos en la penuria económica y en graves dificultades personales. Se argumenta que el FSV fue la causa de muchas de estas clasificaciones incorrectas relacionadas con el fraude.

Pese a que los responsables del Gobierno pidieron disculpas por el escándalo y destinaron 500 millones de euros para compensar a los progenitores afectados en marzo del 2020, el Gobierno de Rutte dimitió a principios de enero del 2021, para evitar perder una moción de confianza en un debate parlamentario.[46]

[41] Blauw 2020

[42] Vijlbrief y van Huffelen 2020a

[43] Vijlbrief y van Huffelen 2020b

[44] KPMG 2020

[45] NL Times 2020

[46] BBC News 2021

Modelos de determinación de guetos para niños en riesgo

En **Dinamarca**, algunas autoridades locales llevaron a cabo un experimento para intentar localizar a niños pequeños vulnerables a raíz de sus circunstancias sociales. El **modelo Gladsaxe**, llamado así por el municipio situado a las afueras de Copenhague donde se inició el proyecto, utilizó un sistema de aprendizaje automático que combinaba información externa con datos de diversas fuentes relacionados con el desempleo, la asistencia sanitaria y las condiciones sociales, para analizar más de 200 indicadores de riesgo.

Dicho modelo empleaba un sistema de puntos, con parámetros como tener una enfermedad mental (3.000 puntos), sufrir desempleo (500 puntos) o haberse ausentado de una cita con el médico (1.000 puntos) o el dentista (300 puntos). El divorcio también se incluyó en la estimación del riesgo, que luego se extendió a todas las familias con hijos, para ayudar a identificar situaciones de vulnerabilidad social. El modelo sumaba o restaba puntos a las familias en función de los datos que se encontraban en el sistema. Esto permitía identificar a niños en riesgo de sufrir abusos o malos tratos para realizar una intervención temprana, que podía dar lugar a traslados forzosos.[47]

El proyecto provocó una reacción social considerable, con quejas de organizaciones de la sociedad civil y de miembros del entorno académico. Quienes lo criticaban se quejaban de que la adopción de una administración algorítmica debilita la rendición de cuentas de las Administraciones, les permite consolidar su poder y conduce inevitablemente a medidas cada vez más draconianas que vigilan el comportamiento individual. En la práctica, se consideraba que este sistema de IA suponía una amenaza para la democracia liberal, y se comparó con el sistema de crédito social empleado por el Gobierno chino.[48]

En un primer momento, el Gobierno danés restó importancia a las críticas, destacó la oportunidad que ofrecía el modelo Gladsaxe para identificar antes a los niños en situación de riesgo y planeó su implantación en todo el país. Esto formaba parte de un "plan de guetos" más amplio para luchar contra las "sociedades paralelas", iniciado en el 2010. El plan del Gobierno incluía el uso de conjuntos de criterios, que iban cambiando, para publicar "listas de guetos" anuales, que definían las zonas donde se consideraba que se concentraban los problemas sociales.[49]

En esas zonas, se aplicarían disposiciones legales especiales, relativas a la prevención de la delincuencia, la integración, la protección de datos, la asistencia social y la asignación de viviendas públicas. Por ejemplo, una iniciativa del 2018 estableció la obligación legal de que los niños que viven en determinados barrios asistieran al menos a 25 horas de guardería obligatoria cada semana a partir de los doce meses. La misma iniciativa también permitía duplicar las sanciones penales en las zonas consideradas guetos.[50]

[47] Thapa 2019

[48] Mchangama y Hin-Yan 2018

[49] El hecho de que el Gobierno usara de manera oficial un término tan cargado de historia como *gueto* ha provocado que las listas de guetos danesas sean objeto de numerosos debates, tanto dentro como fuera de Dinamarca. Fuente: Bendixen 2018

[50] Seemann 2020

Sin embargo, al desvelarse el esquema que utilizaba el modelo Gladsaxe para evaluar el bienestar y el desarrollo de los niños, se supo que las evaluaciones individuales se elaboraban y almacenaban sin el conocimiento de los progenitores y en contra de la legislación vigente. En septiembre del 2018, el ministro responsable mencionó que había una actualización jurídica prevista, pero en diciembre de ese mismo año la propuesta de ampliación del modelo Gladsaxe quedó suspendida, pese a que algunos políticos siguen apoyando que se reinstaure el sistema —con ciertas correcciones— en el futuro.[51]

Perfiles algorítmicos: el nuevo techo de cristal

Al igual que en el caso del sistema de empleo polaco, analizado en la tercera sección, el del **Arbeitsmarktservice (AMS) austríaco**, conocido como *algoritmo AMS*, es otro ejemplo de servicios públicos de empleo (SPE) que utilizan modelos de perfiles algorítmicos para predecir la probabilidad de que los demandantes de empleo encuentren trabajo, en un intento de reducir costes y mejorar la eficiencia. El AMS automatiza la creación de perfiles de los solicitantes de empleo para hacer más eficiente su proceso de asesoramiento y mejorar la eficacia de los planes activos de empleo. A partir de las estadísticas de años anteriores, el sistema calcula las posibilidades futuras de los demandantes de empleo en el mercado laboral mediante el indicador de "posibilidad de reinserción" (valor IC), generado por ordenador.

En la práctica, el sistema algorítmico busca conexiones entre el logro de un empleo y las características de cada demandante, como la edad, el origen étnico, el género, la educación, las tareas de cuidados que realiza y sus problemas de salud, así como el empleo anterior, los contactos con el AMS y la situación del mercado laboral en el lugar de residencia del o la demandante. A continuación, clasifica a los solicitantes de empleo en tres grupos en función de su valor IC previsto: los que tienen grandes posibilidades de encontrar un empleo en seis meses, los que seguramente lo lograrán en un año y los que es probable que consigan un empleo en el plazo de dos años. Posteriormente, se ponen a disposición de las diversas categorías de demandantes diferentes niveles de ayuda y recursos para la formación continua, con el objetivo de invertir sobre todo en aquellos para los que las medidas de apoyo tienen más probabilidades de conducir a la reintegración en el mercado laboral.[52]

El algoritmo fue muy criticado por organizaciones de la sociedad civil, periodistas y académicos, e incluso el *Volksanwaltschaft* (defensor del pueblo), de carácter independiente, expresó preocupación sobre su aplicación. Las críticas tenían que ver con diversos elementos discriminatorios percibidos en el algoritmo, en concreto, respecto a las mujeres y los mayores de 50 años. Aunque el algoritmo se hizo público parcialmente (solo estaban disponibles 2 de las 96 variaciones del modelo), también fue criticado por otros aspectos importantes, como la falta de transparencia o los sesgos del sistema, y por disminuir la capacidad de los trabajadores sociales para tomar decisiones independientes.[53, 54]

[51] Algorithm Watch y Bertelsmann Stiftung 2020

[52] Allhutter et al. 2020

[53] Wimmer 2018

[54] Allhutter et al. 2020

De hecho, aunque el sistema del AMS se había diseñado únicamente para proporcionar al personal una función adicional en la atención a los solicitantes de empleo, un estudio reciente demuestra que tuvo consecuencias de gran alcance para todo el organismo. Estas consecuencias incluían una mayor eficacia del proceso de asesoramiento, pero solo cuando se asociaba a una adopción predominantemente rutinaria del sistema de IA, y una mejora de la "eficacia de la formación" al concentrar la financiación en el medio de los tres grupos. Por otro lado, se confirmó que "al desarrollar el sistema, apenas se aplicaron procedimientos para evitar los sesgos, y el sistema no ofrece indicaciones en su aplicación para prevenir posibles desigualdades estructurales en el trato", en particular en lo que respecta a la igualdad de género.[55]

Como muestran este ejemplo y el anterior de Polonia, estos métodos estadísticos se usan para segmentar a los demandantes de empleo en diferentes grupos, a fin de tratar de identificar mejor a los que corren el riesgo de convertirse en desempleados de larga duración. Pero, al mismo tiempo, inducen a la discriminación. Los sistemas de predicción reflejan sesgos institucionales y sistémicos, y dado que se basan en decisiones de contratación y evaluaciones anteriores, pueden revelar y reproducir patrones de desigualdad, penalizando a los grupos desfavorecidos y minoritarios, incluidas las mujeres.[56]

El ordenador dice que no: efectos en los procedimientos de servicios sociales

La protección social es otro ámbito importante en el que las Administraciones están experimentando con la IA. Entre los ejemplos que están surgiendo en muchos países del mundo, el **caso Trelleborg** merece especial atención.[57, 58] En el año 2016, el **municipio sueco** de Trelleborg comenzó a utilizar un sistema automatizado de toma de decisiones específico, basado en la automatización robótica de procesos (ARP), para gestionar solicitudes de asistencia social, como la asistencia domiciliaria y las prestaciones por enfermedad y desempleo.

La ARP es un proceso que se rige por un sistema de reglas diseñadas por expertos, y su objetivo es automatizar tareas administrativas rutinarias, como el cálculo de tarifas y prestaciones de la asistencia domiciliaria; a continuación, un gestor de casos de la ARP debe confirmar los resultados. En la práctica, sin embargo, el programa informático suele basarse en diferentes reglas que conducen a una decisión (aprobación o denegación), y el gestor de casos suele seguir el criterio del programa.[59]

[55] Institute of Technology Assessment of the Austrian Academy of Sciences 2020

[56] Digital Future Society 2020a

[57] Misuraca y van Noordt 2020

[58] Engstrom et al. 2020

[59] Algorithm Watch y Bertelsmann Stiftung 2020

Para implementar la ARP, fue necesario estructurar y modificar los datos internos y los datos sobre los solicitantes, así como analizar y rediseñar los procesos administrativos. Este ejemplo muestra cómo la IA, si se aplica junto a un proceso de transformación digital, puede mejorar las operaciones de la Administración.[60] El Ayuntamiento de la ciudad argumenta que, de hecho, se ha reducido considerablemente el número de personas que reciben prestaciones sociales de forma incorrecta y que, con el desarrollo futuro del programa, este aprenderá a realizar tareas más complejas y se ampliará así el alcance de la automatización de procesos dentro del sector público.[61]

Dado el aparente éxito del programa, la Agencia Nacional de Innovación de Suecia, Vinnova, y la Asociación Sueca de Autoridades Locales y Regiones colaboraron con Trelleborg para reproducir dicho sistema en otros municipios. Pero el plan sufrió diversas resistencias. Desde el principio, muchos trabajadores sociales temieron perder sus puestos de trabajo, algo comprensible si consideramos que el número de trabajadores sociales se redujo de 11 a 3, y les inquietaba el hecho de dejar tareas sociales tan delicadas en manos de los ordenadores. Otros ayuntamientos suecos que pretendían seguir el ejemplo de Trelleborg también encontraron oposición, y algunos funcionarios dimitieron.

Los informes de casos mencionaron la gran necesidad de que el proceso de automatización fuera fiable. Al tratar de aumentar la eficiencia, hasta un 15 % de las decisiones del sistema eran incorrectas (lo que corresponde a 500.000 casos). Esto provocó el cierre del sistema y numerosas protestas por el riesgo de excluir a ciudadanos vulnerables, ya que con la ARP es más difícil evaluar las necesidades individuales.[62]

En la práctica, como demuestran otros casos de uso de sistemas de IA para automatizar las decisiones sobre prestaciones sociales, la existencia de procesos de documentación tanto informatizados como en papel puede dar lugar a duplicidades e ineficiencias.[63] Además, la falta de confianza en el uso de la IA obliga al personal a revisar de nuevo todos los procesos, lo que de hecho aumenta el tiempo que requiere el servicio y reduce la eficacia.[64]

Calificar o no calificar: perjuicios en Bachillerato

En el 2020, la pandemia de la COVID-19 afectó enormemente a los sistemas educativos de todo el mundo. Dada la gravedad de la situación, el Gobierno del **Reino Unido** decidió eliminar los exámenes de los alumnos de entre 16 y 18 años. Como alternativa, el organismo regulador de exámenes del país desarrolló el **sistema algorítmico de calificación Ofqual**. El objetivo era encontrar una forma objetiva de estandarizar las calificaciones finales de todos los estudiantes, ya que Ofqual había determinado que era injusto evaluarlos basándose únicamente en las valoraciones de los profesores, dadas las diferencias entre unos centros educativos y otros.[65]

[60] Codagnone et al. 2020

[61] UIPath, (s.f.).

[62] Wills 2019

[63] Ranerup y Zinner Henriksen 2019

[64] Wihlborg et al. 2016

[65] BBC News 2020a

Por ello, el sistema de IA combinó las calificaciones anteriores con la evaluación de los profesores, para evitar que se inflaran las notas de algunos alumnos y mantener una distribución adecuada.[66]

El 13 de agosto del 2020, miles de estudiantes del Reino Unido recibieron las calificaciones de sus exámenes de A-level (un nivel similar al Bachillerato). Casi el 40 % de ellos recibieron notas más bajas de lo previsto según las evaluaciones de sus profesores, y el 3 % descendió dos calificaciones respecto de lo esperado.[67] Esto desencadenó protestas sociales y acciones legales. La decisión de optimizar el algoritmo para mantener el mismo nivel y evitar que se inflaran las calificaciones tuvo otras consecuencias inesperadas. Se criticó especialmente el hecho de que el algoritmo rebajaba de forma sistemática los resultados de los alumnos de centros públicos, mientras que mejoraba las notas de los estudiantes de escuelas financiadas con fondos privados. Debido a cómo trataba el algoritmo los grupos pequeños, estaba perjudicando a los alumnos de entornos socioeconómicos más bajos.[68]

En la práctica, los alumnos más brillantes y prometedores de los centros con peores resultados tenían muchas más probabilidades de que les bajara las notas, lo que reducía sus opciones de acceder a las carreras universitarias que querían.[69] En Escocia, por ejemplo, el índice de aprobados de los estudiantes del sistema Higher procedentes de los grupos más desfavorecidos se redujo un 15,2 %, mientras que, en los alumnos de entornos más ricos, ese índice solo disminuyó un 6,9 %.[70]

Ante las numerosas críticas, el Gobierno anunció que los resultados se cambiarían por las estimaciones originales de los docentes. Además, de cara a los exámenes del año siguiente, anunció una consulta pública para recabar opiniones sobre la propuesta de que las calificaciones se determinen según las valoraciones de los profesores sobre el nivel de rendimiento de los alumnos a lo largo del año. La debacle de los exámenes en el Reino Unido ilustra claramente la preocupación, muy real, que existe sobre cuándo o cómo garantizar la legitimidad al utilizar la IA para tomar decisiones que tendrán un gran impacto en las oportunidades vitales de los ciudadanos.

[66] Taylor 2020

[67] Education Technology 2020

[68] Lee 2020

[69] The Conversation 2020

[70] BBC News 2020b

5. Lecciones aprendidas: hay que alejarse de los futuros distópicos

Ha quedado claro que los sistemas de IA, pese a sus avanzadas capacidades y su reputación, un tanto mítica, se enfrentan a los problemas del mundo real cuando se emplean como herramientas inteligentes, seguras y eficientes para ayudar en la toma de decisiones de la Administración y la prestación de servicios públicos.[71]

Por supuesto, no se puede responsabilizar únicamente a la IA de los sesgos y errores asociados a los escándalos que se han señalado en las secciones anteriores. Pero los riesgos que conlleva depender en gran medida de las máquinas ponen de manifiesto, por ejemplo —como señala Philip Alston—, el fracaso sistémico de algunas Administraciones a la hora de proteger a las familias vulnerables frente a unos inspectores fiscales especialmente recelosos, "generando, a todos los niveles, errores que han provocado grandes injusticias para miles de familias y criminalizando a personas inocentes"[72]. Los estudios de casos demuestran el desequilibrio entre los intereses económicos del Estado (la lucha contra el fraude) y el interés social de la privacidad, como confirmaron los tribunales holandeses en el **caso SyRi**.

La gran esperanza de que la IA sea una tecnología inocua, intrínsecamente más transparente, responsable y justa que la toma de decisiones humana, también se ha puesto en duda. Cabe destacar la falta de transparencia y de medidas de protección para garantizar los derechos individuales, como se observó en el **caso Gladsaxe**, por ejemplo. Esto es especialmente relevante al debatir si determinadas situaciones justifican que se recopilen y combinen datos personales, y en qué medida, ya sea para velar por el bienestar de los niños o para combatir una pandemia. En esos casos, el uso de la IA concuerda con los principios de seguridad y protección arraigados en las sociedades, así como con los valores que las sustentan.

De hecho, "los modelos son opiniones integradas en las matemáticas", como explica la científica de datos Cathy O'Neil. "Pese a su reputación de imparcialidad, reflejan los objetivos y la ideología de los seres humanos"[73]. Los modelos son útiles porque nos permiten eliminar la información superflua y centrarnos solo en lo más importante para obtener los resultados que buscamos. Pero también son abstracciones. Las decisiones sobre lo que contienen reflejan las prioridades de sus creadores. Un ejemplo evidente es el **algoritmo AMS**, que representa la transformación hacia un "Estado habilitador"[74] con una transición hacia los regímenes de activación, de manera que la asistencia social deja de ser algo accesible por una cuestión de derechos y pasa a ser un servicio orientado a los consumidores.[75]

[71] Clasen 2021

[72] Alston 2020

[73] O'Neil 2016

[74] Deeming y Smyth 2015

[75] Penz et al. 2017

La naturaleza política inherente a la IA también puede observarse en el **caso Trelleborg**. El sistema de IA implantado en dicha localidad mejoró notablemente un proceso administrativo concreto, pero no pudo garantizar la interoperabilidad de la institución ni ganarse la confianza de la población, ni siquiera la del personal interno. Preocuparon aspectos como el riesgo de excluir a los ciudadanos vulnerables y la "pérdida de control" que supone automatizar todos los procesos.[76]

Esto demuestra la importancia de comprender tanto los retos relacionados con la recopilación y el análisis de datos como los peligros potenciales derivados del diseño de servicios públicos proactivos. El reto se acentúa aún más debido al posible efecto de estigmatización que conlleva clasificar a una persona como un futuro problema en una fase temprana, como se vio en el **caso Gladsaxe**. Y, hablando del futuro, los escandalosos resultados presentados por el **algoritmo de calificación de Ofqual** ilustran los riesgos de poner en manos de los sistemas de IA el control de decisiones cruciales que afectan a la vida de los ciudadanos.

Pero ¿significa esto que los algoritmos nunca podrán llegar al nivel esperado o tomar decisiones?

[76] Codagnone et al. 2020

6. A qué debemos prestar atención

No hay duda de que el despliegue de la IA en el sector público ofrece un enorme potencial para mejorar la vida de los ciudadanos. Por desgracia, y como se ha demostrado en este libro blanco, no es una cuestión sencilla. Al contrario: si la IA no se implementa lo suficientemente bien, reproducirá los sesgos y las limitaciones actuales del ser humano y, a medida que sus usos se vuelvan más sofisticados, esos sistemas pasarán cada vez más desapercibidos, con el potencial de causar graves daños sociales.

Tomemos el ejemplo de los **sistemas de reconocimiento facial (SRF)**, utilizados por millones de personas a diario para iniciar sesión en el móvil, organizar sus fotos o proteger sus dispositivos. Los SRF tienen muchos otros usos beneficiosos aparte de los orientados a los consumidores, como facilitar la vida a las personas invidentes o de baja visión, o ayudar a los organismos policiales a localizar a menores desaparecidos y víctimas de trata de personas.

Sin embargo, pese a la promesa de estos supuestos beneficios, en los últimos dos años ha surgido una notable resistencia a estas tecnologías biométricas, debido a los riesgos para la privacidad, la protección de datos y los derechos humanos que plantea su uso indiscriminado.[77] En todo el mundo, diversos casos de implementación ilegal de SRF han llamado la atención de las organizaciones de derechos digitales y de la ciudadanía en general. Por ejemplo, numerosas ciudades han optado por prohibir que las fuerzas policiales empleen este tipo de tecnología, por temor a que abra la puerta a posibles vulneraciones de la privacidad y a la vigilancia masiva.[78, 79]

En algunos casos, también se ha prohibido realizar pruebas con tecnologías de reconocimiento facial para identificar a posibles delincuentes en lugares públicos, como en el aeropuerto de Zaventem (Bruselas).[80] Asimismo, las peticiones de experimentar con el uso de SRF en centros educativos han recibido críticas y dictámenes negativos en Francia y Suecia. Y, en diversos países, hay un debate sobre la implementación de cámaras corporales para la vigilancia policial.[81]

En el Reino Unido y Francia se han puesto a prueba métodos similares. Por ejemplo, la **Policía Metropolitana de Londres** utilizó dos cámaras de reconocimiento facial en la estación de King's Cross, uno de los lugares más concurridos de la ciudad. El experimento duró meses, y las autoridades no se preocuparon por la transparencia ni pensaron en ofrecer mecanismos de información a los transeúntes cuyos datos habían recogido.[82]

[77] Moraes et al. 2020

[78] Gershgorn 2020

[79] Roussi 2020

[80] Misuraca y van Noordt 2020

[81] Misuraca et al. 2020

[82] Togawa y Deeks 2018

Por su parte, y en el contexto de las medidas de seguridad aplicadas por la COVID-19, la **administración del Metro de París** probó a utilizar la IA para detectar si los viajeros llevaban mascarilla, analizando las imágenes de las cámaras de videovigilancia. La iniciativa formaba parte de los esfuerzos de la ciudad por prevenir la propagación del virus, pero, después de probarse durante tres meses en la céntrica estación de Châtelet-Les Halles, el organismo de protección de datos emitió una advertencia al respecto. Por dicha estación suelen transitar unos 33 millones de pasajeros al año.[83]

La Commission Nationale de l'Informatique et des Libertés (CNIL) argumentó que este tipo de tecnología conlleva el riesgo de que se pueda reconstruir la identidad de las personas analizadas, y que las medidas también estarían sujetas al RGPD, dado que las cámaras recogerían datos personales sin consentimiento.[84]

El mensaje que se desprende de este análisis es claro. Tal y como se explica en el número de febrero de la revista *MIT Technology Review*, titulado muy elocuentemente "This is how we lost control of our faces!" ('¡Así es como perdimos el control de nuestros rostros!'), esta tecnología no solo ha erosionado nuestra privacidad, sino que también ha "alimentado una herramienta de vigilancia cada vez más poderosa. La última generación de reconocimiento facial basado en el aprendizaje profundo [también] ha trastocado por completo nuestras normas relativas al consentimiento"[85].

Los resultados de un estudio reciente, así como un análisis de la mayor encuesta de la historia sobre los SRF, con más de 100 conjuntos de datos de rostros compilados entre los años 1976 y 2019, que contienen 145 millones de imágenes de unos 17 millones de personas, ofrecen varias ideas interesantes. Sugieren que las tecnologías avanzadas de reconocimiento afectan profundamente a la "intimidad" individual, y que ello influirá en cómo las distintas facetas de la sociedad respetan la privacidad, al igual que ha influido en la evolución de esta en los últimos 30 años.[86]

Esto da una idea de cómo evolucionarán los parámetros que definen el uso de los SRF en los próximos 30 años y en los posteriores. Como subrayan los autores en las conclusiones, "los SRF plantean complejos retos éticos y técnicos. De no desentrañar esta complejidad, medirla, analizarla y explicársela a los demás, haríamos un flaco favor a los más afectados por su aplicación negligente"[87].

Pero la IA no se limita a los datos: hay muchos más factores que contribuyen a la innovación basada en la IA. Además de garantizar la disponibilidad de datos de alta calidad para el desarrollo y la adopción de la IA, es crucial asegurarse de que su implementación se ajuste al ámbito y los valores organizativos del sector público, así como a los requisitos específicos que debe cumplir la IA.[88]

[83] Vincent 2020

[84] Fouquet 2020

[85] Hao 2021

[86] Raji y Fried 2021

[87] Ibid.

[88] Misuraca y Viscusi 2020

Para ello, se proponen diferentes opciones de políticas, teniendo en cuenta, por ejemplo, los enfoques basados en la ética desde el diseño, la evaluación previa de la conformidad o la convergencia normativa, y el desarrollo de una contratación pública innovadora.[89, 90]

Además, la confianza de la ciudadanía es esencial para garantizar que estos sistemas sean legítimos y eficaces, sobre todo en el ámbito del sector público. El rápido crecimiento de la literatura en este campo, que muestra los desafíos únicos que presenta el uso de la IA en la Administración, confirma la importancia de la confianza de la ciudadanía, al igual que la atención que le prestan numerosas instituciones, como AI Watch de la Comisión Europea y el Observatorio de Políticas de IA de la OCDE.[91, 92, 93, 94]

Teniendo en cuenta que el desarrollo de la IA se basa en la "combinación de inmensas cantidades de datos con una potente computación y sofisticados modelos matemáticos", una regulación positiva, como describen Gruson et al., debería estudiar detenidamente, y tratar de abordar, los riesgos que suponen la inexactitud y la falta de transparencia.[95] Por lo tanto, es necesario que haya medidas de protección como mecanismos no vinculantes, supervisión, normas internacionales y espacios de pruebas para la regulación.

Tanto en Estados Unidos como en la UE, hay movimientos que abogan por enfoques de regulación específicos, y cada uno de ellos opta por una vía diferente. Aunque la existencia de estas vías en China y en otros países no democráticos no está clara, el número de países afines está creciendo, y el grupo comparte la necesidad de encontrar un enfoque común para desarrollar una IA responsable y centrada en las personas.[96]

En este sentido, un ejemplo interesante al que se debe prestar atención es el **experimento de IA realizado en Espoo (Finlandia)**, que pretendía desarrollar una segmentación de riesgos sociales y sanitarios basándose en evidencias. El objetivo era predecir la trayectoria que seguirían las personas en el futuro, a lo largo de los diversos servicios, para poder aplicar nuevas modalidades de prevención y atención proactiva. El experimento, iniciado en el marco de la Estrategia de las Seis Ciudades para poner a prueba las "sociedades del futuro" en Finlandia, empleó más de 37 millones de datos de interacciones sociosanitarias de unos 520.000 residentes.[97] El sistema integró estos datos con los datos sobre la educación que recibieron en la infancia todos los ciudadanos entre los años 2002 y 2016, y con datos de servicios sanitarios privados y estadísticas nacionales sobre protección social básica.

Este sistema, aunque se considera un éxito, ha quedado en suspenso por el momento, para permitir que se debata sobre la preocupación ética relacionada con el papel del sector público en el desarrollo de estos sistemas y la necesidad de contar con la confianza de los ciudadanos, así como la forma de combinar los distintos conjuntos de datos protegiendo la privacidad y la seguridad.

[89] Foro Económico Mundial 2020

[90] AI Council 2021

[91] Desouza et al. 2020

[92] Sun y Medaglia 2019

[93] Comisión Europea, (s.f.)

[94] Berryhill et al. 2019

[95] Gruson et al. 2019

[96] Feijóo et al. 2020

[97] Engels et al. 2019

Al mismo tiempo, y con la perspectiva opuesta, será interesante observar el posible sucesor de SyRi, cuyo objetivo es luchar contra la delincuencia subversiva y que los críticos ya llaman jocosamente **SuperSyRi**.

Esto confirma que no basta con estar atentos a los aspectos tecnológicos de la IA, como la calidad y la precisión de los datos o la transparencia de los algoritmos, sino que también es necesario generar confianza en esta tecnología disruptiva. Para ello son fundamentales los algoritmos éticos y seguros por su diseño, pero también se necesita que la sociedad civil se involucre más en relación con los valores que deben incorporarse a la IA y las direcciones que deben tomar los futuros desarrollos.[98]

[98] Ada Lovelace Institute 2020

7. Recomendaciones

La IA, si se aplica sabiamente, puede abordar algunos de los retos más inextricables del mundo. Pero los probables efectos desestabilizadores que puede tener en numerosos aspectos de la vida económica y social frustran la importancia de sus repercusiones positivas.[99] Los diversos dilemas a los que se enfrentan los responsables políticos requieren más investigación, debido a las implicaciones imprevistas y los efectos secundarios que pueden tener. A continuación se presentan siete recomendaciones que se deben tener en cuenta en este sentido.

Tener cuidado con el tecnosolucionismo

En primer lugar, hay que evitar pensar en la IA como una especie de superagente capaz de prácticamente todo. Al confiar en métodos automatizados, se reproduce un patrón demasiado familiar: los interesados consideran inicialmente que los sistemas que ayudan a tomar decisiones son fiables y, luego, tras observar errores, desconfían incluso de las aplicaciones más fiables. Adoptar aplicaciones defectuosas demasiado pronto pone en riesgo la confianza en el sistema. Del mismo modo, confiar en las buenas prácticas voluntarias y en la autorregulación es una solución imperfecta, ya que los resultados dependen de la buena fe de actores como Facebook y otras entidades que procesan datos.[100]

También hay que tener en cuenta la percepción de los ciudadanos sobre el hecho de que se compartan sus datos, que puede variar según el contexto cultural y administrativo, y hay que garantizar la posibilidad de incluir contenidos locales, para que se tengan en cuenta diferentes perspectivas.

Desconfiar de los atajos éticos

Al mismo tiempo, debemos ser conscientes de que las tecnologías basadas en la IA, si se gestionan de manera superficial, pueden infringir los principios de privacidad y protección de datos hasta el punto de que no se puedan justificar sus beneficios en materia de seguridad colectiva o calidad de los servicios públicos. Por lo tanto, es importante mantener el nexo entre la consideración de los riesgos éticos y los posibles daños a la cohesión social, por un lado, y las ventajas en términos de eficiencia o productividad que ofrece la IA a un organismo o agencia gubernamental, por otro.

[99] Comisión Europea 2018b

[100] Kuziemski y Misuraca 2020

Asimismo, es esencial examinar detenidamente las barreras que podrían impedir su uso en el sector público, incluidos los efectos no intencionados o inesperados, así como los beneficios potenciales, y comparar los efectos previstos y los observados tras su aplicación. Hay que centrarse en los aspectos jurídicos, técnicos y organizativos, pero también en la aceptación por parte de los ciudadanos.

Basarse en evidencias concretas

Las acciones de numerosas Administraciones en todo el mundo demuestran el creciente interés por explorar y experimentar con el uso de la IA para rediseñar procesos internos del sector público, potenciar mecanismos de elaboración de políticas y mejorar la prestación de servicios públicos. Sin embargo, dado que todavía no hay evidencias claras que confirmen las expectativas depositadas en la IA sobre su supuesto impacto positivo, hay que subrayar el desequilibrio entre la adopción potencial y efectiva de las soluciones de IA.[101]

Además, para abordar debidamente los riesgos éticos y políticos del uso de la IA en el sector público, es primordial la convergencia normativa hacia un enfoque común sobre la adopción de la IA. Esta debería contemplar la reutilización y el uso compartido de sistemas y soluciones basados en IA para los servicios públicos, y la participación de agentes del mundo académico, el sector privado y la sociedad civil en el diseño de los sistemas de IA, así como el ensayo de soluciones alternativas y la evaluación previa tanto de los requisitos de conformidad como de los efectos.

Adoptar una perspectiva de valor público

Si se adopta una perspectiva de valor público, centrada en aplicar la IA de manera eficaz tanto en la Administración como en la prestación de servicios, se abordarán los complejos retos asociados al uso de la IA en las instituciones. De hecho, es fundamental tener en cuenta que, al usar la IA, estamos tratando con "objetos de frontera", un concepto utilizado en sociología para describir fenómenos que "tienen significados diferentes en distintos mundos sociales, pero cuya estructura es lo suficientemente común a más de un mundo como para que sean medios de traducción reconocibles"[102].

En la práctica, las razones para introducir la IA y la percepción de los resultados obtenidos son diferentes para los diversos grupos de interesados. Mientras que, para algunos, el rendimiento y la precisión son los factores más importantes, para otros la trazabilidad, la transparencia y las opciones de subsanación son fundamentales. Lo mismo ocurre con las definiciones individuales de la "calidad" de los servicios, en relación con los datos o la satisfacción de los ciudadanos, por ejemplo.

[101] Misuraca y van Noordt 2020

[102] Star y Griesemer 1989

Prepararse para gestionar la disrupción

Mientras se experimenta con diferentes tecnologías de IA en diversos ámbitos políticos, es importante tener en cuenta el concepto de "reencuadre de la innovación en el sector público", es decir, "la necesidad de considerar tanto los cambios tangibles en los procedimientos, las funciones y las instituciones como una 'reestructuración cognitiva' que concierne a los valores, la cultura y los entendimientos comunes, a fin de articular un conjunto de valores reforzado para la ética del sector público"[103].

Este metaencuadre es necesario al tratar con dinámicas sociales complejas y posiblemente disruptivas e impredecibles, como la IA, para evaluar mejor las consecuencias directas e indirectas de una acción sobre las instituciones, los ciudadanos y la sociedad en conjunto.[104] En última instancia, ello implicará también la necesidad de replantear la forma en que se diseñan y prestan los servicios, el modo en que se comparten y gestionan los datos y la manera en que se aplica la toma de decisiones algorítmica.

Buscar alianzas con las partes interesadas

Reconocer y apreciar las diferentes opiniones y los distintos niveles de comprensión que existen sobre la IA en los grupos clave de la sociedad es fundamental para el éxito de iniciativas complejas, como la adopción de la IA en el sector público. Esto implica llevar a cabo análisis interdisciplinarios y entablar comunicación e interacciones con las diversas partes interesadas, en paralelo a la transformación del sector público. En este contexto, puede ser útil considerar los efectos potenciales de la IA en la consecución de los Objetivos de Desarrollo Sostenible fijados en la Agenda 2030 de la ONU. Esto garantizaría una IA fiable y centrada en las personas, y se aprovecharía su potencial de mejorar el bienestar de todos.[105, 106]

Diseñar nuevos modelos de gobernanza

La gobernanza es un concepto relevante para la IA en tres sentidos. En primer lugar, usar la IA permite al sector público obtener unas ventajas sin precedentes y, en segundo lugar, abre la puerta a la capacidad de influir en los ciudadanos para que se comporten de una manera u otra, con la condición de garantizar un equilibrio adecuado entre la privacidad personal y los derechos humanos. Esto requiere un compromiso con la gobernanza de la IA, que garantice que la IA genere valor público, sea beneficiosa para todos y no se considere solo un objetivo en sí misma.

[103] Misuraca et al. 2020

[104] Rossel 2010

[105] Vinuesa et al. 2020

[106] Feijóo et al. 2020

Por último, es necesario aprender a gobernar el uso de la IA en el sector público para vincularlo progresivamente al impacto, más amplio, que puede tener en diversos ámbitos políticos. Aunque son pocas las ocasiones en que se ha implementado correctamente, es crucial identificar y compartir estudios de casos de uso para aprender sobre la IA, replicarla, ampliarla e institucionalizarla en los servicios generales.[107] Solo así superaremos el *impasse* de experimentar eternamente y no llegar nunca a instaurar lo que realmente funciona, al tiempo que desterraremos para siempre las verdaderas amenazas que ponen en riesgo la estabilidad de nuestras sociedades.

[107] Misuraca y van Noordt 2020

Referencias

Ada Lovelace Institute. (2020). Examining the Black Box: Tools for assessing algorithmic systems. [online] Disponible en: https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/

AI Council. (2021). AI Roadmap. Office for Artificial Intelligence, Department for Business, Energy & Industrial Strategy, and Department for Digital, Culture, Media & Sport. GOV.UK. [online] Disponible en: https://www.gov.uk/government/publications/ai-roadmap

Algorithm Watch y Bertelsmann Stiftung. (2020) Automating Society Report 2020. [online] Disponible en: https://automatingsociety.algorithmwatch.org

Allhutter, D., Cech, F., Fischer, F., Grill, G. y Mager, A. (2020). Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective. Frontiers in Big Data. [online] Disponible en: https://www.frontiersin.org/article/10.3389/fdata.2020.00005

Alston, P. (2019). Digital technology, social protection and human rights: Report. Naciones Unidas. [online] Disponible en: https://www.ohchr.org/EN/Issues/Poverty/Pages/DigitalTechnology.aspx

Alston, P. (2020). Landmark ruling by Dutch court stops government attempts to spy on the poor. Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos. [online] Disponible en: https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?LangID=E&NewsID=25522

Babuta, A. y Oswald, M. (2019). Data Analytics and Algorithmic Bias in Policing, Briefing paper, Royal United Services Institute for Defence and Security Studies. UK government's Centre for Data Ethics and Innovation. [PDF] Disponible en: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/831750/RUSI_Report_-_Algorithms_and_Bias_in_Policing.pdf

Barenstein, M. (2019). ProPublica's COMPAS Data Revisited. Cornell University. [online] Disponible en: https://arxiv.org/abs/1906.04711v3

BBC News. (2020a). A-levels: Why are students so unhappy about this year's results? [online] Disponible en: https://www.bbc.co.uk/newsround/53803651

BBC News. (2020b). Scotland's results day: Thousands of pupils have exam grades lowered. [online] Disponible en: https://www.bbc.com/news/uk-scotland-53636296

BBC News. (2021). Dutch Rutte government resigns over child welfare fraud scandal. [online] Disponible en: https://www.bbc.com/news/world-europe-55674146

Bendixen, M. (2018). Denmark's 'anti-ghetto' laws are a betrayal of our tolerant values. The Guardian. [online] Disponible en: https://www.theguardian.com/commentisfree/2018/jul/10/denmark-ghetto-laws-niqab-circumcision-islamophobic

Berryhill, J., Kok Heang, K., Clogher, R. y McBride, K. (2019). Hello, World! Artificial intelligence and its use in the Public Sector. iLibrary de la OCDE. [online] Disponible en: https://www.oecd.org/governance/innovative-government/working-paper-hello-world-artificial-intelligence-and-its-use-in-the-public-sector.htm

Big Brother Watch. (2020). Big Brother Watch briefing on Algorithmic Decision-Making in the Criminal Justice System. [PDF] Disponible en: https://bigbrotherwatch.org.uk/wp-content/uploads/2020/02/Big-Brother-Watch-Briefing-on-Algorithmic-Decision-Making-in-the-Criminal-Justice-System-February-2020.pdf

Blauw, S. (2020). An algorithm was taken to court – and it lost. The Correspondent. [online] Disponible en: https://thecorrespondent.com/276/an-algorithm-was-taken-to-court-and-it-lost-which-is-great-news-for-the-welfare-state/36504050352-a3002ff7

Bronstein, H. (2020). Rights group criticizes Buenos Aires for using face recognition tech on kids. Reuters. [online] Disponible en: https://www.reuters.com/article/ctech-us-argentina-rights-idCAKBN26U23Z-OCATC

Clasen, S. (2021). When the government uses AI: Algorithms, differences, and trade-offs. ASU. W. P. Carey News. [online] Disponible en: https://news.wpcarey.asu.edu/20210119-when-government-uses-ai-algorithms-differences-and-trade-offs

Codagnone, C., Liva, G., Barcevičius, E., Misuraca, G., Klimavičiūtė, L., Benedetti, M., Vanini, I., Vecchi, G., Ryen Gloinson, E., Stewart, K., Hoorens, S. y Gunashekar, S. (2020). Assessing the impacts of digital government transformation in the EU. Oficina de Publicaciones de la UE. [online] Disponible en: https://op.europa.eu/en/publication-detail/-/publication/7e715248-aac0-11ea-bb7a-01aa75ed71a1/language-en

Comisión Europea. (n.d.). AI for the public sector. [online] Disponible en https://knowledge4policy.ec.europa.eu/ai-watch/topic/ai-public-sector_en

Comisión Europea. (2018a). Artificial Intelligence for Europe. [online] Disponible en: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:237:FIN

Comisión Europea. (2018b). The Age of Artificial Intelligence Towards a European Strategy for Human-Centric Machines. European Political Strategy Centre. [online] Disponible en: https://ec.europa.eu/jrc/communities/sites/jrccties/files/epsc_strategicnote_ai.pdf

Comisión Europea. (2020a). Shaping Europe's digital future. [online] Disponible en: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0067

Comisión Europea. (2020b). White Paper on Artificial Intelligence: a European approach to excellence and trust. [online] Disponible en: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

Craglia, M., Annoni, A., Benczur, P., Bertoldi, P., Delipetrev, P., De Prato, G., Feijoo, C., Fernández Macías, E., Gómez, E., Iglesias, M., Junklewitz, H, López Cobo, M., Martens, B., Nascimento, S., Nativi, S., Polvora, A., Sánchez, I., Tolan, S., Tuomi, I. y Vesnic Alujevic, L. (2018). Artificial Intelligence - A European Perspective. Oficina de Publicaciones. [online] Disponible en: https://www.researchgate.net/publication/329449889_Artificial_Intelligence_A_European_Perspective

Danaher, J., (2016). The Threat of Algocracy: Reality, Resistance and Accommodation. Philosophy & Technology. [online] Disponible en: https://www.scinapse.io/papers/2242985385

Deeming, C. y Smyth, P. (2015). Social Investment after Neoliberalism: Policy Paradigms and Political Platforms. [online] Disponible en: https://www.cambridge.org/core/journals/journal-of-social-policy/article/social-investment-after-neoliberalism-policy-paradigms-and-political-platforms/C8E670BB1F0E2185F0EDDFB4B8C5AB8E

Dencik, L., Hintz, A., Redden, J. y Warne, H. (2018). Data Scores as Governance: Investigating uses of citizen scoring in public services. Open Society Foundations. [PDF] Disponible en: https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf

Dencik, L. Redden, J. Hintz, A. y Warne, H. (2019). The 'golden view': data-driven governance in the scoring society. Internet Policy Review. [online] Disponible en: https://doi.org/10.14763/2019.2.1413

Desouza, K., Dawson, G. y Chenok, D. (2020). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. Business Horizons, 63(2), 205–213. [online] Disponible en: https://doi.org/10.1016/j.bushor.2019.11.004

Digital Future Society. (2020a). Inclusión por diseño: exploración de diseños sensibles al género en el bienestar digital. [online] Disponible en: https://digitalfuturesociety.com/es/report/exploring-gender-responsive-designs-in-digital-welfare/

Digital Future Society. (2020b). Brecha de género en los datos: hacia la igualdad de género en el bienestar digital. [online] Disponible en: https://digitalfuturesociety.com/report/hacia-la-igualdad-de-genero-en-el-estado-de-bienestar-digital/

Douglas Heaven, W. (2020). Predictive policing algorithms are racist. They need to be dismantled. MIT Technology Review. [online] Disponible en: https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

Education Technology. (2020). 36% of A-levels in England downgraded by Ofqual algorithm. [online] Disponible en: https://edtechnology.co.uk/he-and-fe/36-of-a-level-grades-in-england-downgraded-by-ofqual-algorithm

Engels, F., Wentland, A. y Pfotenhauer, S. M. (2019). Testing future societies? Developing a framework for test beds and living labs as instruments of innovation governance. Research Policy. [online] Disponible en: https://www.sciencedirect.com/science/article/pii/S0048733319301465

Engstrom, D. F., Ho, D. E., Sharkey, C. M. y Cuéllar, M. F. (2020). Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies. Revista electrónica SSRN. [online] Disponible en: https://doi.org/10.2139/ssrn.3551505

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. Nueva York, EE. UU: "St. Martin's Press.

European Union Agency for Fundamental Rights. (2019). Facial recognition technology: fundamental rights considerations in the context of law enforcement. [online] Disponible en: https://op.europa.eu/en/publication-detail/-/publication/0de97f99-10db-11ea-8c1f-01aa75ed71a1/language-en

Feijóo, C., Kwon, Y., Bauer, J., M., Bohlin, E., Howell, B., Jain, R., Potgieter, P., Vu, K., Whalley, J. y Xia, J. (2020). Harnessing artificial intelligence to increase wellbeing for all: The case for a new technology diplomacy. Telecommunications Policy. [online] Disponible en: https://doi.org/10.1016/j.telpol.2020.101988

Foro Económico. (2020). The Global Risks Report 2020. World Economic Forum. [PDF] Disponible en: http://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf

Feldstein, S. (2019). The Global Expansion of AI Surveillance. Carnegie Endowment for International Peace. [online] Disponible en: https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847

Fouquet, H. (2020). Paris Tests Face-Mask Recognition Software on Metro Riders. Bloomberg. [online] Disponible en: https://www.bloombergquint.com/politics/paris-tests-face-mask-recognition-software-on-metro-riders

Gershgorn, D. (2020). Live Facial Recognition Is Spreading Around the World. OneZero. Medium. [online] Disponible en: https://onezero.medium.com/live-facial-recognition-is-spreading-around-the-world-13f128c671dc

GOV.UK. (s.f.). Troubled Families Programme. [online] Disponible en: https://troubledfamilies.blog.gov.uk/

Grupo Independiente de Expertos de Alto Nivel Sobre Inteligencia Artificial. (2019). Directrices éticas para una IA fiable. Comisión Europea. [online] Disponible en: https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es

Gruson, D., Helleputte, T., Rousseau, P. y Gruson, D. (2019). Data science, artificial intelligence, and machine learning: Opportunities for laboratory medicine and the value of positive regulation. Clinical Biochemistry. [online] Disponible en: https://pubmed.ncbi.nlm.nih.gov/31022391/

Hao, K. (2021). This is how we lost control of our faces. MIT Technology Review. [online] Disponible en: https://www.technologyreview.com/2021/02/05/1017388/ai-deep-learning-facial-recognition-data-history/

Institute of Technology Assessment of the Austrian Academy of Sciences. (2020). An Algorithm for the unemployed? Socio-technical analysis of the so-called "AMS Algorithm" of the Austrian Public Employment Service (AMS). [online] Disponible en: https://www.oeaw.ac.at/en/ita/projects/ams-algorithm

Kharpal, A. (2017). Stephen Hawking says A.I. could be 'worst event in the history of our civilization'. CNBC. [online] Disponible en: https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html

Kolkman, D. (2020). The usefulness of algorithmic models in policy making. Government Information Quarterly, 37(3), 101488. [online] Disponible en: https://doi.org/10.1016/j.giq.2020.101488

KPMG. (2020). Rapportage verwerking van risicosignalen voor toezicht. KPMG Advisory N.V. [PDF] Disponible en: https://www.rijksoverheid.nl/documenten/kamerstukken/2020/07/10/kpmg-rapport-fsv-onderzoek-belastingdienst

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. [online] Disponible en: https://doi.org/10.1016/j.telpol.2020.101976

Lee, G. (2020). Did England exam system favour private schools? Channel 4 News. [online] Disponible en: https://www.channel4.com/news/factcheck/factcheck-did-england-exam-system-favour-private-schools

Mchangama, J. y Hin-Yan, L. (2018). The Welfare State Is Committing Suicide by Artificial Intelligence. Foreign Policy. [online] Disponible en: https://foreignpolicy.com/2018/12/25/the-welfare-state-is-committing-suicide-by-artificial-intelligence/

Misuraca, G., Barcevičius, E. y Codagnone, C. (2020). Exploring Digital Government Transformation in the EU – Understanding public sector innovation in a data-driven society. Comisión Europea. [online] Disponible en: https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/exploring-digital-government-transformation-eu-understanding-public-sector-innovation-data

Misuraca, G. y van Noordt, C. (2020). AI Watch – Artificial Intelligence in public services: Overview of the use and impact of AI in public services in the EU. Comisión Europea. [online] Disponible en: https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/ai-watch-artificial-intelligence-public-services

Misuraca, G. y Viscusi, G. (2020). AI-Enabled Innovation in the Public Sector: A Framework for Digital Governance and Resilience. International Conference on Electronic Government. EGOV 2020. [online] Disponible en: https://link.springer.com/chapter/10.1007%2F978-3-030-57599-1_9

Moraes, T. G., Almeida, E. C. y de Pereira, J. R. L. (2020). Smile, you are being identified! Risks and measures for the use of facial recognition in (semi-)public spaces. AI Ethics. [online] Disponible en: https://doi.org/10.1007/s43681-020-00014-3

Morozov, E. (2020). The tech 'solutions' for coronavirus take the surveillance state to the next level. The Guardian. [online] Disponible en: https://www.theguardian.com/commentisfree/2020/apr/15/tech-coronavirus-surveilance-state-digital-disrupt

NL Times. (2020). Parents faced 'unprecedented injustice' for years in childcare subsidy scandal. [online] Disponible en: https://nltimes.nl/2020/12/17/parents-faced-unprecedented-injustice-years-childcare-subsidy-scandal

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Nueva York, EE. UU.: Crown Publishing Group.

Oxford Insights. (2020). AI Readiness Index 2020. [online] Disponible en: https://www.oxfordinsights.com/government-ai-readiness-index-2020

Parlamento Europeo. (2016). Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) (Texto pertinente a efectos del EEE). [online]. Disponible en: https://eur-lex.europa.eu/eli/reg/2016/679/oj

Penz, O., Sauer, B., Gaitsch, M., Hofbauer, J. y Glinsner B. (2017). Post-bureaucratic encounters: Affective labour in public employment services. Critical Social Policy. [online] Disponible en: https://doi.org/10.1177/0261018316681286

Raji I., y Fried, G. (2021). About Face: A Survey of Facial Recognition Evaluation. [PDF] Disponible en: https://arxiv.org/pdf/2102.00813.pdf

Ranerup, A. y Zinner Henriksen, H. (2019). Value positions viewed through the lens of automated decision-making: The case of social services. Government Information Quarterly 36, 101377. [online] Disponible en: https://doi.org/10.1016/j.giq.2019.05.004

Rechtbank Den Haag. (2020). SyRI legislation in breach of European Convention on Human Rights. de Rechtspraak. [online] Disponible en: https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx

Richardson, R., Schultz, J. y Crawford, K. (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. 94 N.Y.U. L. Rev. Online 192. [online] Disponible en: https://ssrn.com/abstract=3333423

Rossel, P. (2010). Making anticipatory systems more robust. Foresight. [online] Disponible en: https://doi.org/10.1108/14636681011049893

Roussi, A. (2020). Resisting the rise of facial recognition. Nature. [online] Disponible en: https://www.nature.com/articles/d41586-020-03188-2

Seemann, A. (2020). The Danish 'ghetto initiatives' and the changing nature of social citizenship, 2004–2018. Critical Social Policy. [online] Disponible en: https://doi.org/10.1177/0261018320978504

Star, S. L. y Griesemer, J. (1989). Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology. Social Studies of Science, Vol. 19, p. 387-420.

Sun, T. Q. y Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly 36. [online] Disponible en: https://doi.org/10.1016/j.giq.2018.09.008

Taylor, R. (2020). Written statement from Chair of Ofqual to the Education Select Committee. GOV.UK. Ofqual. [online] Disponible en: https://www.gov.uk/government/news/written-statement-from-chair-of-ofqual-to-the-education-select-committee

Thapa, E. P. (2019). Predictive Analytics and AI in Governance: Data-driven government in a free society – Artificial Intelligence, Big Data and Algorithmic Decision-Making in government from a liberal perspective. European Liberal Forum. [PDF] Disponible en: https://liberalforum.eu/publication/predictive-analytics-and-ai-in-governance-data-driven-government-in-a-free-society/

The Conversation. (2020). Gavin Williamson, Ofqual and the great A-level blame game. [online] Disponible en: https://theconversation.com/gavin-williamson-ofqual-and-the-great-a-level-blame-game-144766

Togawa Mercer, S. y Deeks, A. (2018). 'One Nation Under CCTV': The U.K. Tackles Facial Recognition Technology. Lawfare blog. [online] Disponible en: https://www.lawfareblog.com/one-nation-under-cctv-uk-tackles-facial-recognition-technology

UIPath. (s.f.). RPA in the Public Sector: UiPath Helps Swedish Citizens Regain Self-Sufficiency. [online] Disponible en: https://www.uipath.com/resources/automation-case-studies/trelleborg-municipality-enterprise-rpa

Vijlbrief, J. A. y van Huffelen, A. C. (2020a). Informatie over de Fraude Signalering Voorziening (FSV) en het gebruik van FSV binnen de Belastingdienst. Tweede Kamer Der Staten-General. [online] Disponible en: https://www.tweedekamer.nl/kamerstukken/detail?id=2020Z13850&did=2020D29414

Vijlbrief, J. A. y van Huffelen, A. C. (2020b). Kamerbrief Fraude Signalering Voorziening (FSV). Ministrie van Financien. [PDF] Disponible en: https://www.rijksoverheid.nl/documenten/kamerstukken/2020/04/28/kamerbrief-fraude-signalering-voorziening-fsvKamerbrief+Fraude+Signalering+Voorziening+%28FSV%29.pdf

Vincent, J. (2020). France is using AI to check whether people are wearing masks on public transport. The Verge. [online] Disponible en: https://www.theverge.com/2020/5/7/21250357/france-masks-public-transport-mandatory-ai-surveillance-camera-software

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Daniela Langhans, S., Tegmark, M. y Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications. [online] Disponible en: https://doi.org/10.1038/s41467-019-14108-y

Wihlborg, E., Larsson, H. y Hedström, K. (2016). "The Computer Says No!" – A Case Study on Automated Decision-Making in Public Authorities. 49th Hawaii International Conference on System Sciences (HICSS). [online] Disponible en: https://ieeexplore.ieee.org/document/7427547

Wills, T. (2019). Sweden: Rogue algorithm stops welfare payments for up to 70,000 unemployed. Algorithm Watch. [online] Disponible en: https://algorithmwatch.org/en/rogue-algorithm-in-sweden-stops-welfare-payments/

Wimmer, B. (2018). Der AMS-Algorithmus ist ein „Paradebeispiel für Diskriminierung". Kurier – futurezone. [online] Disponible en: https://futurezone.at/netzpolitik/der-ams-algorithmus-ist-ein-paradebeispiel-fuer-diskriminierung/400147421

Agradecimientos

Autor principal

- **Gianluca Misuraca**, vicepresidente de Inspiring Futures

Autora secundaria

- **Tanya Álvarez**, investigadora de Digital Future Society Think Tank

Equipo de Digital Future Society Think Tank

- **Carina Lopes**, directora de Digital Future Society Think Tank
- **Patrick Devaney**, editor de Digital Future Society Think Tank
- **Olivia Blanchard**, investigadora de Digital Future Society Think Tank

Citas

Este informe se debe citar de la siguiente manera:

- Digital Future Society. (2021). Gobernanza y algoritmos: riesgos y potencial del uso de la inteligencia artificial en el sector público. Barcelona, España.

Datos de contacto

Si desea ponerse en contacto con el equipo de Digital Future Society Think Tank, envíe un correo electrónico a thinktank@digitalfuturesociety.com


Digital
Future Society