

Guía práctica ¿Cómo abrir los algoritmos públicos?

Recomendaciones para la implantación de registros de algoritmos públicos

Versión 1.0 Junio 2024



Coordinación: Lorenzo Cotino Hueso

Autores:

José Manuel Calabuig, *Catedrático Universidad. Director Instituto Universitario de Matemática Pura y Aplicada-UPV.*

Lorenzo Cotino, *Catedrático Derecho Constitucional, Universitat de València, Valgrai.*

Antonia Ferrer Sapena, *Catedrática de Universidad, Directora del Observatorio de Datos Abiertos y Transparencia*

Enrique Alfonso Sánchez Pérez, *Catedrático de Universidad, Dpto. de Matemática Aplicada, Instituto Universitario de Matemática Pura y Aplicada*

Publicada la versión 1.0 en junio 2024 por el Observatorio de Transparencia y Datos Abiertos de la Universitat Politècnica de València (UPV) con la Cátedra PAGODA Universitat de València, Instituto Interuniversitario de Desarrollo Local.

Esta publicación ha sido realizada con el soporte económico de la Generalitat Valenciana. Presidència.

El contenido de la publicación es responsabilidad exclusiva de la Universitat Politècnica de València y de la Universitat de València y no representa necesariamente a la Generalitat Valenciana.

© Calabuig, J. M., Cotino, L., Ferrer, A. y Sánchez E. A

Cita sugerida

Calabuig, J. M., Cotino, L., Ferrer, A. y Sánchez E. A, *Guía práctica ¿Cómo abrir los algoritmos públicos? Recomendaciones para la implantación de registros de algoritmos públicos*, Cotino Hueso, Lorenzo (coord), Universitat Politècnica de València- Universitat de València, Versión 1.0 Junio 2024.

ISBN: 978-84-123363-5-1

Esta publicación está disponible en acceso abierto bajo la licencia Attribution-ShareAlike 3.0 IGO (CCBY-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo/>).

CONTENIDO GENERAL

I. ¿Por qué es necesaria la transparencia de los algoritmos públicos?	4
<i>¿Para qué se usan los algoritmos y la IA en el sector público?</i>	4
<i>Casos polémicos por falta de transparencia</i>	9
<i>El escándalo de la Oficina de Correos británica y las lecciones para la transparencia los algoritmos públicos</i>	17
II. A qué hay que dar transparencia pública, no sólo sistemas de IA, también automatizados y algoritmos	20
III. Más riesgo, más transparencia. ¿Cómo evaluar el riesgo de los algoritmos públicos?	21
<i>Más riesgo, más transparencia según el impacto en las personas y sus derechos</i>	21
<i>¿Cómo evaluar el riesgo de los algoritmos públicos?</i>	23
<i>Criterios específicos para la evaluación del riesgo de los algoritmos</i>	31
<i>Escala general de riesgo de los algoritmos según las funciones que desarrollan</i>	31
<i>¿Cómo valorar del uno al cinco el riesgo global de los sistemas integrales de IA en el sector público?</i>	38
IV. ¿Qué información debe aportarse para el registro de algoritmos de IA?	39
<i>Propuesta de ficha de registro</i>	40
V. Un “paseo” por registros de algoritmos públicos y mejores prácticas	42
<i>Algunos estándares y registros destacados</i>	42
<i>Una valoración de estos estándares y experiencias</i>	50
VI. Para acabar ¿Cómo implementar un registro de algoritmos públicos?	53
Anexos	56
<i>Anexo I. Sistemas del Ayuntamiento de Barcelona en su primera guía</i>	56
<i>Anexo II. Referencia a algunos algoritmos públicos</i>	59
<i>Anexo III. Ficha extraída de Radar: “Radar de algoritmos y procesos de decisión automatizada para el acceso a los derechos sociales de la ciudadanía”</i>	61
<i>Anexo IV. La Guía de información algorítmica para el ámbito laboral</i>	67
<i>Anexo V. Ficha de transparencia algorítmica de la AOC en Cataluña</i>	70
<i>Anexo VI. Tabla Resumen comparado de la información proporcionada en los registros o estándares establecidos o propuestos</i>	72

I. ¿POR QUÉ ES NECESARIA LA TRANSPARENCIA DE LOS ALGORITMOS PÚBLICOS?

¿PARA QUÉ SE USAN LOS ALGORITMOS Y LA IA EN EL SECTOR PÚBLICO?

No es muy fácil saber para qué se está usando la IA en el sector público, puesto que muchos usos son secretos o como mínimo discretos. Por el contrario, los usos que se conocen, por lo general, es porque la propia administración quiere que se conozcan y en muchas ocasiones puede pensarse que son los que mejor quedan a ojos de la ciudadanía.



Imagen creada con Dall-E

En todo caso, a partir de los datos y fuentes disponibles, es posible afirmar que la IA y los algoritmos públicos se usan para:

- **Mejorar la calidad de los servicios públicos**

Según la OCDE, uno de los principales objetivos del uso de IA es mejorar la calidad de los servicios de bienestar social. Por ejemplo, los chatbots y asistentes virtuales se utilizan para interactuar con los ciudadanos, proporcionando respuestas rápidas y eficientes a sus consultas. Estos sistemas son especialmente útiles en el

ámbito de la salud, donde pueden ayudar a gestionar citas médicas y brindar información sobre tratamientos.

- **Aumentar la eficiencia operativa**

La automatización de tareas repetitivas y que consumen mucho tiempo es una de las principales ventajas de la IA. Por ejemplo, en el sector de los recursos humanos, los sistemas de IA pueden ayudar a evaluar solicitudes de empleo y gestionar la nómina, liberando al personal para que se enfoque en tareas más estratégicas.

- **Tomar decisiones basadas en datos**

La IA permite analizar grandes volúmenes de datos y extraer información útil para la toma de decisiones. En el ámbito de la salud pública, la IA se utiliza para monitorear y analizar riesgos, como brotes de enfermedades, y para predecir la demanda de servicios médicos. Un ejemplo es el sistema utilizado en Estonia para predecir lluvias y su impacto en la tierra, lo que ayuda a tomar decisiones informadas sobre la gestión de recursos naturales.

- **Garantizar la seguridad y el orden público**

Los sistemas de reconocimiento inteligente, como los sistemas biométricos y de videovigilancia, son comunes en este ámbito. Por ejemplo, en los Países Bajos, se utilizan sistemas de IA para predecir accidentes de tráfico y mejorar la seguridad vial. En Francia, los servicios de aduanas emplean IA para detectar declaraciones de valores erróneos.

- **Personalización de servicios**

La IA también se utiliza para personalizar los servicios públicos según las necesidades individuales de los ciudadanos. Un ejemplo destacado es el uso de chatbots en el sector de la salud para ofrecer recomendaciones personalizadas sobre tratamientos y cuidados preventivos. En el ámbito de la protección social, estos sistemas ayudan a gestionar las solicitudes de beneficios y a ofrecer asistencia específica según las necesidades de cada persona.

- **Análisis y predicción en la educación**

En el sector educativo, la IA se usa para analizar datos y predecir tendencias. Por ejemplo, en Portugal, se utiliza un sistema de IA para planificar el flujo de estudiantes en el sistema educativo, y en Suecia, se emplea para detectar las causas del abandono escolar prematuro.

- **Protección ambiental**

La IA desempeña un papel crucial en la preservación del medio ambiente. Sistemas avanzados de reconocimiento digital se utilizan para identificar especies de árboles y monitorear la tala de bosques. En Italia, un sistema de IA predice la lluvia y su impacto en la tierra, lo que ayuda a gestionar mejor los recursos naturales y a prevenir desastres ambientales.

- **Configuración de estrategias y liderazgo tecnológico**

Los gobiernos desarrollan estrategias de IA para mostrar compromiso político y avanzar en la agenda del gobierno digital. Este enfoque permite a las administraciones mantenerse a la vanguardia en tecnologías emergentes y garantizar que las políticas públicas sean modernas y efectivas. Por ejemplo, muchos gobiernos están implementando marcos regulatorios y políticas para guiar el desarrollo y uso de IA de manera ética y segura.

- **Financiación**

Los gobiernos financian con fondos directos o indirectos para apoyar la investigación, desarrollo y adopción de tecnologías emergentes. Esto incluye la financiación de proyectos de investigación en universidades y centros de investigación, así como la concesión de subvenciones a empresas que desarrollan soluciones innovadoras en IA. Este apoyo financiero es crucial para fomentar la innovación y asegurar que las nuevas tecnologías se integren efectivamente en los servicios públicos.

- **Las Administración como consumidores directos de IA**

Las administraciones públicas también son usuarios directos de tecnologías de IA, adquiriendo y comprando servicios innovadores o co-desarrollando soluciones a medida. Un ejemplo de esto es el uso de sistemas de IA para mejorar la eficiencia de los procesos administrativos, como la gestión de recursos humanos o la planificación urbana inteligente. Al adoptar estas tecnologías, los gobiernos pueden ofrecer servicios más rápidos y eficientes a los ciudadanos.

- **Coordinador y regulador**

Además de ser usuarios y financiadores, los gobiernos actúan como coordinadores y reguladores, facilitando la cooperación entre diferentes actores y estableciendo normas y políticas tanto a nivel nacional como internacional. Esto asegura que el uso de IA esté alineado con los valores y objetivos sociales, y que se minimicen los riesgos asociados con su implementación.

- **Hacer cumplir mandatos regulatorios**

La IA se utiliza para mejorar la eficiencia en el cumplimiento de mandatos regulatorios, incluyendo la seguridad en el trabajo, el cuidado de la salud y la protección ambiental. Por ejemplo, los sistemas de IA pueden monitorear las condiciones de trabajo y detectar posibles riesgos para la seguridad, asegurando que las empresas cumplan con las normativas vigentes.

- **Adjudicación de subvenciones y ayudas**

Los algoritmos de IA también se utilizan en la adjudicación de subvenciones y ayudas, desde beneficios por discapacidad hasta derechos de propiedad intelectual. Estos sistemas pueden evaluar solicitudes de manera rápida y precisa, garantizando que los recursos se distribuyan de manera justa y eficiente.

- **Control y análisis de riesgos**

En áreas como la salud y la seguridad pública, la IA se utiliza para el monitoreo y análisis de riesgos. Por ejemplo, los sistemas de IA pueden analizar datos de salud pública para identificar brotes de enfermedades y predecir su propagación, lo que permite a las autoridades tomar medidas preventivas de manera oportuna.

- **Extracción de información de datos masivos**

La IA facilita la extracción de información útil de grandes volúmenes de datos, como las quejas de consumidores o los patrones climáticos. Esta capacidad es fundamental para que los gobiernos puedan tomar decisiones informadas y diseñar políticas efectivas basadas en datos reales.

- **Comunicación con el público**

Finalmente, la IA se utiliza para mejorar la comunicación con el público, informando a los ciudadanos sobre sus derechos y obligaciones. Por ejemplo, los sistemas de IA pueden enviar notificaciones personalizadas a los beneficiarios de asistencia social, contribuyentes, solicitantes de asilo y dueños de negocios, asegurando que reciban la información relevante de manera oportuna y efectiva.

Fuentes: Hay que destacar de inicio el mejor estudio existente sobre casos de uso IA públicos, si en 2020 (JRC, 2020) fue de 240 casos de uso, son más de 600 en 2022 (JRC, Tangi, 2022) para la UE, entre los tipos de IA que se emplean en el sector público. También desde la OCDE (2019) se señalan diversos papeles de los gobiernos con la IA. Freeman y otros (2020: 9 y 21 y ss.) para EEUU.

JRC, Misuraca, G. (2020): *AI Watch. Artificial Intelligence in public services. Overview of the use and impact of AI in public services in the EU*, Joint Research Centre, Unión Europea.

JRC, Tangi, L. y otros (2022): *AI Watch European landscape on the use of Artificial Intelligence by the Public Sector*, JRC Science For Policy Report, Unión Europea.

OCDE, Barbara Ubaldi y otros (2019b): *State of the art in the use of emerging technologies in the public sector*, OECD Working Papers on Public Governance No. 31.

Freeman Engstrom, D. y otros (2020): "Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies. Report submitted to the Administrative conference of the United States", NYU School of Law, Public Law Research Paper No. 20-54, 122 págs.

dl.acm.org - Introduction to the Issue on Artificial Intelligence in the Public Sector

- snowflake.com - Use Cases of AI in Government
- adalovelaceinstitute.org - Public sector use of data and algorithms
- research.aimultiple.com - AI in Government: Examples, Challenges & Best Practices
- dl.acm.org - Introduction to the Issue on Artificial Intelligence in the Public Sector
- v7labs.com - 7 Practical Applications of AI in Government
- snowflake.com - Use Cases of AI in Government
- adalovelaceinstitute.org - Public sector use of data and algorithms
- hellotars.com - Top 10 Applications of Artificial Intelligence in Government
- oecd-ilibrary.org - Casos prácticos de uso de la IA en los Gobiernos
- unesco.org - Inteligencia artificial: ejemplos de dilemas éticos
- noticias.juridicas.com - Un nuevo impulso al uso de la Inteligencia Artificial en la gestión pública
- redalyc.org - Inteligencia Artificial en la administración pública
- idpbarcelona.net – Cinco claves para entender la transparencia algorítmica en el sector público

CASOS POLÉMICOS POR FALTA DE TRANSPARENCIA

La aplicación de la IA en la administración pública ha sido altamente cuestionada por *falta de transparencia en los algoritmos*. La apertura en los algoritmos permitiría detectar fallos y mejorar la equidad, pero las preocupaciones de seguridad nacional y propiedad intelectual a menudo impiden esta transparencia.

Además de las finalidades, destacamos a continuación algunos casos que finalmente han resultado polémicos, muchas veces por la falta de transparencia del sistema de IA público.



Imagen creada con runwayml

ÁREA DE AYUDAS PÚBLICAS

BOSCO (España): Utilizado para el bono social eléctrico. Los fallos detectados por la Fundación Civio motivaron la solicitud del código fuente de la aplicación. La opacidad del sistema levantó dudas sobre su equidad y funcionamiento.

Fundación Civio - Bono Social Eléctrico: <https://civio.es/el-boe-nuestro-de-cada-dia/2020/02/06/el-deficit-de-transparencia-del-bono-social-electrico/>



Servicios sociales

Modelo Gladsaxe (Dinamarca): Este sistema, empleado por el municipio de Gladsaxe, identificaba y evaluaba a niños en riesgo de abuso. Utilizaba algoritmos que integraban información de salud de los padres, desempleo y datos locales. Fue eliminado en 2019 tras críticas sobre privacidad y alcance ético, generadas por la falta de transparencia en su funcionamiento.

The Guardian - Denmark's 'predictive policing' of children:
<https://www.theguardian.com/society/2019/nov/18/child-protection-ai-predict-prevent-risks>

Is it right to use AI to identify children at risk of harm?

Machine learning is being used to help protect children, but it raises ethical questions



Allegheny (EE.UU.): Sistema de IA para el seguimiento del bienestar infantil, cuestionado y con intervenciones judiciales debido a problemas en su implementación, amplificados por la falta de claridad en cómo se tomaban las decisiones.

The Verge - Inside Allegheny County's controversial child welfare algorithm: <https://www.theverge.com/2020/8/15/21370411/allegheny-county-predictive-algorithm-child-welfare>

ÁMBITO POLICIAL Y SISTEMA PENAL

Informe junio 2024: Universidad de Valencia con Amnistía Internacional: Tres enfoques policiales predictivos en España: Viogén, RisCanvi y Veripol (una evaluación desde una perspectiva de derechos humanos), de Martínez Garay et al. https://www.uv.es/regulation/papers/MartinezGaray2024_Predictive_policing_Spain.pdf



Prisiones: RisCanvi (España) y COMPAS (EE.UU.): Estiman el riesgo de reincidencia de delincuentes. Estas herramientas han sido criticadas por posibles sesgos y falta de transparencia en sus algoritmos.

Un algoritmo define el futuro de los presos en cataluña: ahora sabemos cómo funciona: https://www.elconfidencial.com/tecnologia/2024-04-24/riscanvi-algoritmo-cataluna-prisiones-presos-inteligencia-artificial_3871170/

Vera Institute - Understanding COMPAS: <https://www.vera.org/publications/understanding-compas>

El País - RisCanvi: el sistema catalán para predecir reincidencias: https://elpais.com/ccaa/2019/04/23/catalunya/1556027794_672337.html



VioGén (España): Sistema de seguimiento integral para casos de violencia de género, predice el riesgo de ataque tras analizar 37 indicadores. La falta de claridad en los criterios algorítmicos ha suscitado preocupaciones.

La Vanguardia - Así funciona VioGén, el sistema que protege a las víctimas de violencia de género:
<https://www.lavanguardia.com/vida/20181125/453118859391/viogen-sistema-proteccion-violencia-genero.html>

VeriPol (España): Detecta denuncias falsas relacionadas con robos y hurtos, utilizando combinaciones frecuentes en casos de falsedad. La opacidad en los criterios de detección ha sido motivo de crítica.

Policía.es
https://www.policia.es/es/comunicacion_prensa_detalle.php?ID=4433&idiomaActual=es#



PredPol (EE.UU.): Algoritmo de predicción delictiva basado en el historial de delitos. Criticado por discriminación racial y enfoque en zonas de pobreza, sin transparencia en su diseño y operativa.

¿Cómo funciona Predpol, el software que dice predecir dónde van a suceder crímenes? <https://www.xataka.com/aplicaciones/como-funciona-predpol-el-software-que-dice-predecir-donde-van-a-suceder-crimes>

EDUCACIÓN

Ofqual (Reino Unido): Durante la pandemia de 2020, se utilizó un algoritmo para asignar calificaciones universitarias, lo que resultó en una reducción general de notas y protestas que llevaron a su retirada. La falta de transparencia en los criterios del algoritmo fue un punto crítico.

Ofqual: <https://www.gov.uk/government/publications/ofquals-approach-to-regulating-the-use-of-artificial-intelligence-in-the-qualifications-sector/ofquals-approach-to-regulating-the-use-of-artificial-intelligence-in-the-qualifications-sector#:~:text=move%20beyond%20these.-,Ofqual's%20approach%20to%20regulating%20AI,or%20public%20confidence%20in%2C%20qualifications>. BBC - A-levels and GCSEs: How did the exam algorithm work?: <https://www.bbc.com/news/explainers-53807730>



Policy paper

Ofqual's approach to regulating the use of artificial intelligence in the qualifications sector

Published 24 April 2024

Applies to England

EMPLEO

AMS (Austria): Sistema de IA para asignar puntuaciones a desempleados. Se descubrió que discriminaba a las mujeres, asignándoles peores puntuaciones que a hombres con igual experiencia y habilidades. Fue declarado ilegal y suspendido por la autoridad de protección de datos, debido en parte a la opacidad del sistema.

Reuters - Austria halts controversial job-matching algorithm:
<https://www.reuters.com/article/us-austria-jobs-algorithm-idUSKBN2A20IG>

DETECCIÓN DEL FRAUDE

SyRI (Países Bajos): Diseñado para combatir el fraude en programas sociales y fiscales. Fue suspendido en 2020 por el tribunal de La Haya debido a preocupaciones sobre la privacidad y la falta de transparencia.

The Guardian - Dutch court rules against government's use of 'racial' algorithm to fight fraud: <https://www.theguardian.com/technology/2020/feb/05/dutch-court-rules-against-governments-use-of-racial-algorithm-to-fight-fraud>

Toeslagenaffaire (Países Bajos): Otro sistema para la detección de fraude, también polémico por la opacidad de sus criterios y resultados.

DutchNews - Dutch childcare benefits scandal: how it happened:
<https://www.dutchnews.nl/news/2021/01/dutch-childcare-benefits-scandal-how-it-happened/>

Robodebt (Australia): Utilizado para reclamar deudas a ciudadanos que reciben ayudas públicas. El sistema tenía inexactitudes, causando problemas de salud mental y física a los afectados. El estado indemnizó a las víctimas y el sistema fue desactivado. La falta de claridad en el funcionamiento del algoritmo contribuyó a su controversia.

The Guardian - Australia's robodebt scandal explained:
<https://www.theguardian.com/australia-news/2020/may/29/australias-robodebt-scandal-explained>

OTROS CASOS DE USO

Detección de menores vulnerables (Dinamarca): Un proyecto en Gladsaxe fue retirado por la imposición de medidas severas y la negativa a almacenar información sensible. La falta de transparencia en el tratamiento de los datos fue un problema clave.

The Guardian - Denmark's 'predictive policing' of children:
<https://www.theguardian.com/world/2020/feb/11/denmark-predictive-policing-children-welfare>

Condado de Allegheny (EE.UU.): Utilizó un algoritmo para predecir el riesgo de maltrato doméstico en menores, llevando a investigaciones sociales injustas y pérdida de custodia. La falta de claridad en la evaluación algorítmica fue criticada.

The Verge - Inside Allegheny County's controversial child welfare algorithm: <https://www.theverge.com/2020/8/15/21370411/allegheny-county-predictive-algorithm-child-welfare>

EL ESCÁNDALO DE LA OFICINA DE CORREOS BRITÁNICA Y LAS LECCIONES PARA LA TRANSPARENCIA LOS ALGORITMOS PÚBLICOS

El Escándalo de la Oficina de Correos (Post Office Scandal, <https://www.postofficescandal.uk/>) es un ejemplo claro de cómo la falta de transparencia y la confianza ciega en la tecnología pueden llevar a grandes injusticias. Este escándalo involucró el procesamiento y condena equivocada de más de 900 trabajadores postales por robo y fraude entre 1999 y 2015, todo debido a un sistema informático defectuoso llamado Horizon.



Home News Sport Business Innovation Culture Travel Earth Video Live

Post Office Horizon scandal: Why hundreds were wrongly prosecuted

24 May 2024

Share



El Post Office Scandal tuvo graves consecuencias humanas, incluyendo al menos *cuatro suicidios* entre los subadministradores de correo que fueron injustamente acusados y condenados por robo y fraude debido a errores en el sistema Horizon. Además de estos trágicos suicidios, muchos otros enfrentaron la ruina financiera y sufrieron graves impactos psicológicos, como ansiedad, depresión y estrés postraumático.

El coste total del escándalo ha sido extremadamente alto. Hasta el momento, el escándalo ha costado a los contribuyentes británicos más de *£1,000 millones*. Además, el gobierno ha pagado aproximadamente *£160 millones* en compensaciones a más de 2,700 víctimas a través de tres esquemas diferentes. Se espera que Fujitsu, la compañía que desarrolló el sistema Horizon, contribuya significativamente a estos costos, posiblemente hasta un 50% de los mismos.

LA HISTORIA DETRÁS DEL ESCÁNDALO

En 1999, la Oficina de Correos británica implementó Horizon, un software contable que pronto comenzó a reportar posibles faltas de dinero -posibles delitos o infracciones-, que eran incorrectos. En lugar de investigar y corregir los errores del sistema, la gerencia procesó a cientos de subadministradores de correos por fraude y malversación de fondos. *La falta de transparencia* en el sistema y la confianza ciega en la tecnología llevó a *injusticias graves*, incluyendo la ruina financiera, encarcelamientos y al menos cuatro suicidios.

La falta de transparencia y sus consecuencias

La Oficina de Correos insistió durante años en que Horizon era confiable, a pesar de que sabía desde 2003 que el sistema tenía defectos. La falta de transparencia permitió que se siguieran procesando a los trabajadores hasta 2015. *Este encubrimiento y la negativa a reconocer los fallos del sistema subrayan la necesidad de transparencia en los algoritmos utilizados por el sector público.*

Cuando finalmente se estrenó el drama televisivo *Mr Bates vs. The Post Office* en enero de 2024, el gobierno británico se vio obligado a actuar. El primer ministro, Rishi Sunak, anunció una legislación de emergencia para revocar las condenas de 980 subadministradores de correos y ofrecer paquetes de compensación. Paula Vennells, la ex CEO de la Oficina de Correos, devolvió su título de Comandante de la Orden del Imperio Británico y posiblemente devolverá 2,2 millones de libras en bonos.



La importancia de la transparencia en los algoritmos públicos

El Escándalo de la Oficina de Correos demuestra por qué es crucial que los algoritmos utilizados en el sector público sean transparentes y comprensibles. *La falta de transparencia no solo permite errores injustos, sino que también erosiona la confianza pública en las instituciones.* Los sistemas algorítmicos deben ser auditables y los resultados deben poder ser explicados y verificados por terceros independientes.

Lecciones aprendidas

1. Transparencia en los algoritmos públicos: La falta de transparencia en el funcionamiento de Horizon permitió que los errores del sistema pasaran desapercibidos y que las injusticias continuaran durante años. Es fundamental que los algoritmos públicos sean transparentes y auditables para evitar abusos y errores.
2. Responsabilidad de las compañías desarrolladoras: A pesar de ser una compañía reconocida, Fujitsu falló en reconocer y corregir los defectos de Horizon a tiempo. Incluso las mejores y más conocidas compañías deben ser responsables y transparentes en sus desarrollos tecnológicos.
3. Supervisión adecuada: La Oficina de Correos, a pesar de ser una corporación estatal, operaba con mínima supervisión gubernamental, lo que permitió que se tomaran decisiones sin la debida diligencia. Es crucial que haya mecanismos de supervisión robustos para garantizar que los sistemas de IA se utilicen de manera justa y efectiva.
4. Criterios claros y verificables: Los algoritmos deben basarse en criterios claros y verificables, evitando suposiciones incorrectas.
5. No confiar ciegamente en la tecnología: Incluso los desarrollos de las compañías más reconocidas pueden fallar. La supervisión humana es indispensable.
6. Necesidad de intervención temprana: La intervención gubernamental tardía en el escándalo de la Oficina de Correos muestra la importancia de actuar rápidamente ante señales de fallos sistémicos. Las comisiones de revisión y otros mecanismos deben estar bien equipados y ser eficaces para abordar los problemas de manera oportuna.¹

¹ Algunas fuentes:

<https://www.almendron.com/tribuna/el-escandalo-de-la-oficina-de-correos-britanica-y-el-estado-de-derecho/>

[Institute for Government - Six lessons government should learn from the Post Office scandal](#)

[Open Access Government - Lessons learned from Post Office's Horizon IT scandal](#)

[Thomas Murray - Unexpected lessons from the UK Post Office scandal](#)

[Guildhall Chambers - The Post Office: Disclosure Lessons from a National Scandal](#)

[LinkedIn - Lessons from the UK Post Office Scandal](#)

[Dynamic Boards - What Can Non-Executive Directors Learn From The Post Office Horizon Scandal](#)

[Computer Weekly - Post Office Horizon scandal explained: Everything you need to know](#)

[Post Office Scandal - Venal. Incompetent. Mendacious.](#)

[Wikipedia - British Post Office scandal](#)

[Law Gazette - Post Office live: Former head of criminal Jarnail Singh worried about brand reputation](#)

[BBC - Post Office Horizon scandal: Why hundreds were wrongly prosecuted](#)

[YouTube - LIVE: Post Office inquiry: watch as former executives give evidence](#)

II. A QUÉ HAY QUE DAR TRANSPARENCIA PÚBLICA, NO SÓLO SISTEMAS DE IA, TAMBIÉN AUTOMATIZADOS Y ALGORITMOS

El uso de la inteligencia artificial en el sector público puede ser especialmente peligroso debido a su autonomía, opacidad y difícil predictibilidad. Sin embargo, no solo la IA requiere transparencia. El uso público de "simples" sistemas de algoritmos, software o sistemas automatizados también puede ser peligroso y generar impactos. Estos sistemas, aunque menos complejos que la IA, pueden igualmente afectar negativamente a las personas si no se manejan adecuadamente. Un claro ejemplo de esto es el caso que se ha visto del servicio de correo Británico, pues se trataba software sin IA para gestionar las operaciones contables.

La ley valenciana es un buen punto de partida en esta dirección, porque abarca los sistemas de IA, pero también el uso de algoritmos o los sistemas automatizados. Además, esta ley no se limita a las decisiones exclusivamente automatizadas sin intervención humana ni al impacto directo en individuos específicos. Lo mismo sucede en la llamada *ley rider*, según el artículo 64.4.d) del Estatuto de los Trabajadores, los representantes sindicales deben tener información de los algoritmos que afectan la toma de decisiones sobre los trabajadores, incluso si el algoritmo solo apoya la decisión final y, por supuesto, aunque no sea IA

Asimismo, la transparencia no debe limitarse solo al uso de IA o algoritmos en los procedimientos administrativos, sino también a la prestación de servicios públicos. En muchos casos, la IA se utiliza para personalizar servicios, operar chatbots y asistentes, realizar identificación biométrica, gestionar información, y apoyar decisiones en políticas de transporte, sostenibilidad, y planificación gubernamental. Estos usos tienen un impacto significativo en la ciudadanía y sus derechos, por lo que la transparencia en estos casos es igualmente crucial.



III. MÁS RIESGO, MÁS TRANSPARENCIA. ¿CÓMO EVALUAR EL RIESGO DE LOS ALGORITMOS PÚBLICOS?

MÁS RIESGO, MÁS TRANSPARENCIA SEGÚN EL IMPACTO EN LAS PERSONAS Y SUS DERECHOS

La transparencia en el uso de la inteligencia artificial y los algoritmos en el sector público es esencial debido a los diversos impactos y riesgos que estos sistemas pueden tener en los derechos de las personas y otros bienes protegidos.

- Importancia de la transparencia según el impacto
 - La ley valenciana establece que la publicidad de los sistemas debe ser proporcional a su impacto. Esto significa que:
 - Los sistemas sin impacto significativo no requieren publicidad.
 - Los sistemas con mayor impacto necesitan una transparencia más intensa.
 - El impacto debe ser un criterio básico para determinar qué sistemas deben incluirse en el registro de algoritmos.
- Factores que determinan la necesidad de transparencia

Las Directrices para el sector público de Países Bajos y el enfoque del Reino Unido destacan que la transparencia debe depender del impacto del algoritmo en la decisión, el resultado y el ciudadano; el grado de autonomía en la toma de decisiones y el tipo y la complejidad del algoritmo... La necesidad de un registro y la intensidad

de la transparencia depende también de factores como la opacidad y el tipo de algoritmo o el grado de participación, control y supervisión humana en las decisiones automatizadas.

En Barcelona han optado por aplicar los protocolos de actuación a los sistemas de IA de alto riesgo, según lo estipulado por el futuro Reglamento de IA.

A falta de una decisión concreta en la normativa que establezca la transparencia y explicabilidad de la IA, se hace necesario evaluar el impacto de los algoritmos públicos para determinar la transparencia exigible. En los diversos estudios que hemos realizado, consideramos que estos son los parámetros principales que hay que tener en cuenta para fijar el mayor o menor impacto. A continuación, presentamos una serie de preguntas claras e interesantes que se deben realizar respecto del algoritmo, software o sistema del que se trate. Responder a estas preguntas permitirá evaluar el nivel de impacto y riesgo del algoritmo y, en consecuencia, determinar las exigencias de transparencia y explicabilidad necesarias para garantizar su uso justo y responsable en el sector público.

- Factores de riesgo a considerar y preguntas:
 1. Efectos en grupos protegidos:
 - ¿Cómo afecta este algoritmo a los grupos vulnerables o protegidos?
 2. Impactos jurídicos y significativos:
 - ¿Qué consecuencias legales o significativas tiene el uso de este algoritmo?
 3. Personalización de servicios e información:
 - ¿Hasta qué punto este algoritmo personaliza los servicios y la información para los usuarios?
 4. Individualización de decisiones:
 - ¿Este algoritmo toma decisiones individuales que afectan a personas específicas?
 5. Impacto en decisiones políticas o normativas generales:
 - ¿Cómo influye este algoritmo en las decisiones políticas o normativas generales?
 6. Escalabilidad y generalización de errores:
 - ¿Existe el riesgo de que los errores se escalen y generalicen?
 7. Perpetuación de sesgos no controlados:
 - ¿Está perpetuando este algoritmo sesgos que no han sido controlados o corregidos?
 8. Niveles de automatización y menor intervención y supervisión humana:
 - ¿Qué nivel de automatización tiene este sistema? ¿cuánta intervención humana se da? ¿hay supervisión?
 9. Tipo y complejidad del algoritmo:
 - ¿Qué tipo de algoritmo se utiliza y cómo de complejo es?
 10. Opacidad del sistema o algoritmo:
 - ¿Qué grado de opacidad tiene el sistema o algoritmo para la propia Administración y sus supervisores? ¿y para los usuarios?

¿CÓMO EVALUAR EL RIESGO DE LOS ALGORITMOS PÚBLICOS?

¿CÓMO EVALUAR EL RIESGO DE LOS ALGORITMOS SEGÚN SUS FUNCIONES?

Un elemento para la evaluación del riesgo de los algoritmos utilizados en los sistemas de IA depende del tipo de acción que desarrollan. A continuación, se presenta una escala de niveles de riesgo basada en las funciones de estos sistemas.



Imagen creada con Dall-E

Escala general de niveles de riesgo de los algoritmos

1. Sistemas de alto riesgo según el RIA

Descripción: Aquellos sistemas declarados de alto riesgo por el Reglamento de Inteligencia Artificial (RIA) deben ser controlados y cumplir toda una serie de obligaciones.

Ejemplo: Sistemas de identificación biométrica utilizados en aeropuertos para control de fronteras, que requieren altos niveles de precisión y seguridad para evitar errores que puedan poner en riesgo la seguridad de los ciudadanos.



Fuente: propuesta de regulación de la IA. AI Act. Comisión Europea, 2021

2. Sistemas de control y priorización

Descripción: Sistemas de control de aplicación de la ley, de priorización de acciones de inspección fiscal y control del fraude, o sistemas de decisión sobre actuaciones que afecten significativamente la integridad física y jurídica y la salud de individuos particulares.

Ejemplo: Algoritmos utilizados para detectar fraudes fiscales en declaraciones de impuestos, que deben ser precisos para evitar errores que puedan perjudicar a ciudadanos inocentes.

3. Sistemas de adjudicación

Descripción: Sistemas de adjudicación de contratos y subvenciones.

Ejemplo: Algoritmos que asignan contratos públicos a empresas, garantizando la transparencia y evitando favoritismos o corrupción.

4. Sistemas de gestión de servicios ciudadanos

Descripción: Sistemas para la gestión de servicios prestados a los ciudadanos en áreas como salud, vivienda y ayudas sociales.

Ejemplo: Sistemas que asignan citas médicas en hospitales públicos, optimizando los recursos y tiempos de espera para los pacientes.

5. Sistemas de obtención de información

Descripción: Sistemas para la obtención de información utilizada en la elaboración de políticas y la toma de decisiones por parte de la Generalitat Valenciana.

Ejemplo: Algoritmos que analizan datos socioeconómicos para diseñar políticas de desarrollo urbano.

6. Sistemas de gestión interna

Descripción: Sistemas para la gestión interna de procesos y recursos de la administración.

Ejemplo: Sistemas de planificación de recursos humanos dentro de las administraciones públicas, asegurando una distribución eficiente del personal.

7. Sistemas de información e interacción con los ciudadanos

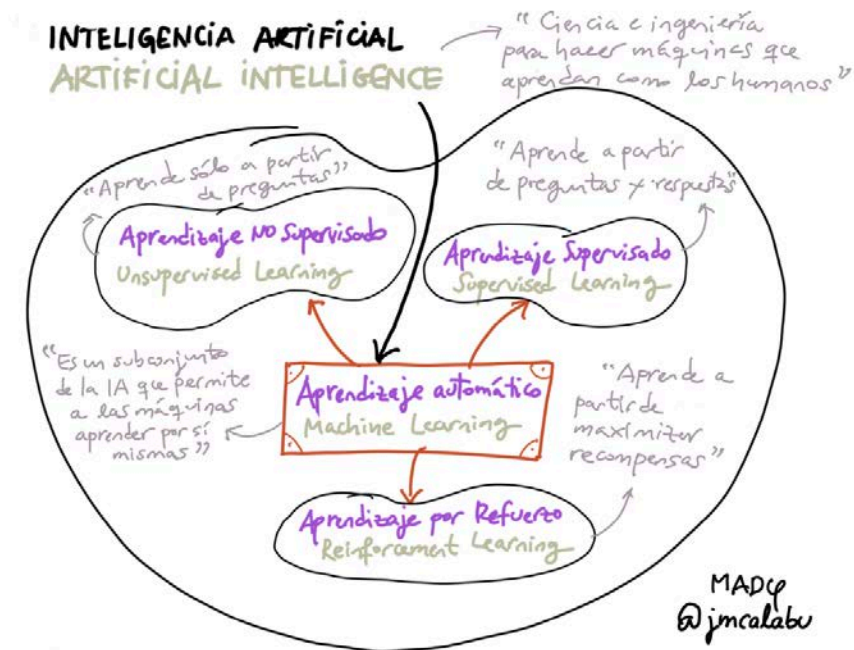
Descripción: Sistemas de información e interacción con los ciudadanos para la prestación de servicios.

Ejemplo: Chatbots utilizados en los sitios web gubernamentales para responder preguntas frecuentes y guiar a los ciudadanos en la realización de trámites administrativos.

¿CÓMO AFECTA EL TIPO DE IA QUE SE UTILIZA?

El aprendizaje automático (*machine learning*) es posiblemente la rama de la IA que más impacto tiene en la generación de algoritmos de utilidad pública. Se divide en tres paradigmas según su diseño general:

- *Aprendizaje supervisado:* Aprende a partir de preguntas y respuestas.
- *Aprendizaje no supervisado:* Aprende a partir solo de preguntas.
- *Aprendizaje reforzado:* Aprende a partir de maximizar recompensas asignadas a las respuestas previas.



A su vez los algoritmos permiten alcanzar diferentes objetivos:

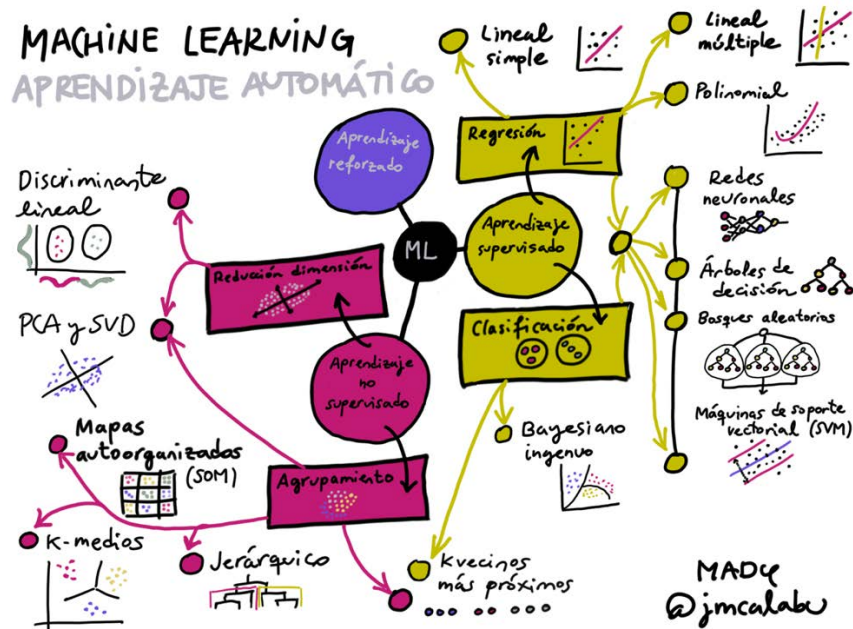
1. **Algoritmos de clasificación:** como su propio nombre indica permiten clasificar "individuos" (personas, documentos, facturas...) en función de unas determinadas características. De esta manera cada "individuo" pertenecerá a un grupo. Algunos de los algoritmos de clasificación más utilizados podrían ser:

- **Bayesiano ingenuo:** Gestión de créditos bancarios. *Ejemplo:* Clasificar si un cliente es apto para recibir un crédito bancario basado en su historial financiero.
- **Árboles de decisión:** Detección de enfermedades. *Ejemplo:* Identificar si un paciente tiene diabetes a partir de sus síntomas y análisis de laboratorio.
- **Bosques aleatorios:** Detección de spam en el correo electrónico. *Ejemplo:* Filtrar correos electrónicos no deseados en tu bandeja de entrada.
- **Máquinas de soporte vectorial:** Detección facial y clasificación de textos. *Ejemplo:* Reconocer caras en fotos subidas a una red social.
- **K-vecinos próximos:** Perfiles de usuarios de servicios. *Ejemplo:* Clasificar a los usuarios de una plataforma de streaming según sus preferencias de visualización.

2. **Algoritmos de agrupamiento:** la diferencia con la clasificación es que en este caso los "individuos" se agrupan por similitud. En este tipo de algoritmos las características de los "individuos" permiten decir qué "individuos" serían similares. Entre otros, algunos métodos utilizados serían:

- **Método k-medias (k-means):** Análisis de sentimiento. *Ejemplo:* Agrupar comentarios de redes sociales en categorías como positivos, negativos y neutros.

- **Agrupamiento jerárquico:** Procesamiento analítico de redes sociales. *Ejemplo:* Analizar las interacciones en una red social para identificar comunidades de usuarios con intereses similares.
3. **Algoritmos de regresión:** este tipo de algoritmos permite, fundamentalmente, hacer predicciones a partir de los datos. Un método de este tipo de algoritmos sería:
- **Regresión supervisada:** Predicción de valores de variables de interés. *Ejemplo:* Estimar los niveles esperados de glucosa en sangre para pacientes diabéticos en diferentes momentos del día.
4. **Combinación de algoritmos:** Muchos sistemas combinan varios algoritmos para aumentar su efectividad. *Ejemplo:* ChatGPT combina modelos de procesamiento de lenguaje natural, técnicas avanzadas de aprendizaje a partir de datos autogenerados y algoritmos de aprendizaje reforzado para entender y generar texto.



IMPORTANCIA DE LOS DATOS EN LOS ALGORITMOS. EL ALIMENTO DE LA IA

Además de los algoritmos en sí, el proceso de adquisición de datos es crucial. Los datos que alimentan los algoritmos pueden contener errores y ser la principal fuente de riesgo. Ejemplo: Un sistema de aprendizaje de peligrosidad basado en datos de barrios de minorías étnicas podría generar un sesgo racista en las decisiones tomadas por el sistema.

El tipo de IA o tecnología utilizada puede no ser decisiva para la evaluación del riesgo de los sistemas algorítmicos. Sin embargo, es cierto que algunos sistemas pueden representar un riesgo potencial más alto. Los sistemas de IA varían en su peligrosidad e impacto dependiendo de sus niveles de automatización, complejidad y opacidad. Los sistemas más automatizados, complejos y opacos presentan mayores riesgos, mientras que aquellos con mayor supervisión humana, menor complejidad y mayor transparencia son menos riesgosos y más fáciles de controlar.



Imagen creada con runwayml

Así, los sistemas de IA que operan con un alto nivel de automatización y mínima supervisión humana son generalmente más peligrosos. Esto se debe a que la falta de intervención humana puede permitir que los errores y los sesgos en los datos se propaguen sin corrección. Por ejemplo, los algoritmos de clasificación como los bosques aleatorios y las máquinas de soporte vectorial, que se utilizan en la detección de spam o reconocimiento facial, respectivamente, pueden operar de manera autónoma, lo que puede llevar a decisiones incorrectas sin intervención humana oportuna.

De lado contrario, serán menos peligrosos los sistemas con menor automatización y mayor supervisión humana. La intervención humana permite la revisión y corrección de

decisiones algorítmicas, reduciendo así el impacto de errores y sesgos. Por ejemplo, los algoritmos de regresión supervisada, utilizados para predecir niveles de glucosa en sangre, requieren intervención humana para ajustar y verificar los resultados.

Por cuanto a la complejidad del algoritmo. Los algoritmos más complejos, como las máquinas de soporte vectorial y los modelos combinados utilizados en sistemas avanzados como ChatGPT, tienden a ser más opacos y difíciles de interpretar. La alta complejidad puede ocultar errores y sesgos, haciendo difícil la supervisión y corrección. Estos sistemas también requieren conocimientos especializados para su correcta implementación y supervisión. Del lado contrario, los algoritmos menos complejos, como los árboles de decisión y los métodos k-medias, son más fáciles de entender y supervisar. Su simplicidad permite una mayor transparencia y facilita la detección y corrección de errores. Por ejemplo, un árbol de decisión utilizado para la detección de enfermedades es más fácil de interpretar y ajustar que un modelo de soporte vectorial.

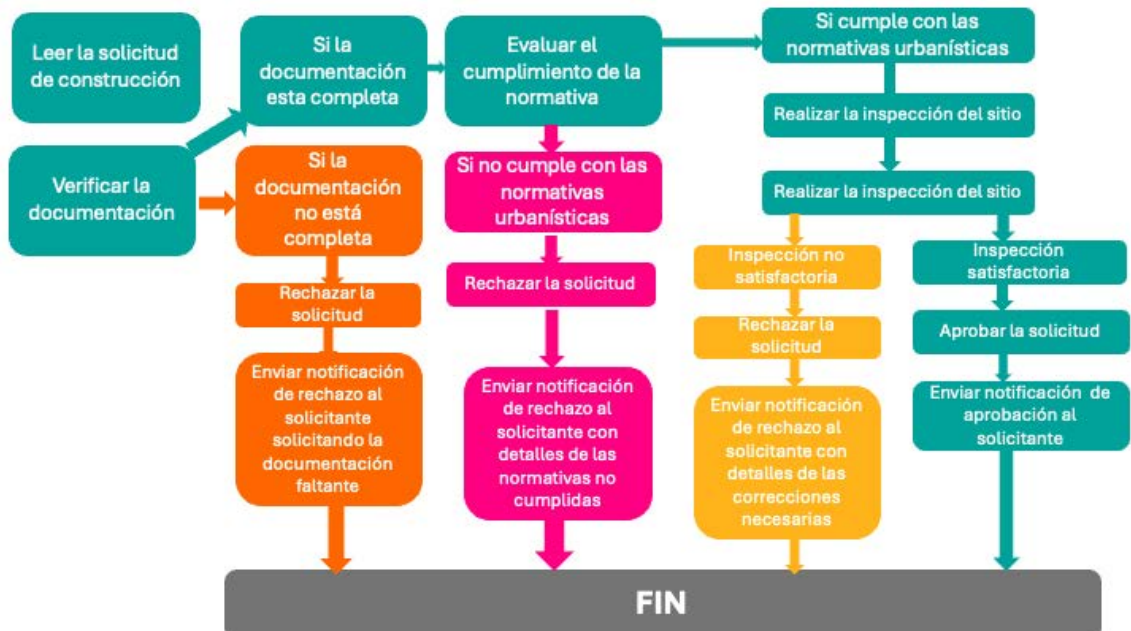
Finalmente, teniendo en cuenta la opacidad del sistema o algoritmo, los sistemas con alta opacidad, como las redes neuronales profundas y algunos algoritmos de aprendizaje no supervisado, son difíciles de interpretar tanto para los usuarios como para los supervisores. Esta falta de transparencia puede ocultar decisiones incorrectas y sesgadas, y hace difícil la auditoría y el control. Un ejemplo es el uso de algoritmos de reconocimiento facial, que pueden tener decisiones opacas y difíciles de explicar. Y al revés. Los sistemas con baja opacidad son aquellos cuyos procesos y decisiones son más transparentes y comprensibles. Los algoritmos como el bayesiano ingenuo y los árboles de decisión suelen ser más transparentes y fáciles de auditar. Esto facilita la comprensión de cómo se toman las decisiones y permite una mayor confianza y supervisión. Por ejemplo, el uso de árboles de decisión para clasificar si un cliente es apto para recibir un crédito bancario es más transparente y fácilmente justificable.

En definitiva, quizás sea necesario sacrificar el rendimiento de los algoritmos para poder obtener sistemas menos opacos. Además, y como ya se hace en ciertas áreas de investigación, una buena práctica que aportaría transparencia sería la publicación del llamado “pseudocódigo” de los algoritmos. El pseudocódigo no es más que una descripción de un algoritmo que utiliza una combinación de lenguaje natural y estructuras de programación. Esta descripción permite detallar los pasos que sigue el algoritmo para proporcionar sus resultados y en este no se hace uso de la sintaxis específica de ningún lenguaje de programación. Por ejemplo:

Inicio

- 1. Leer solicitud de permiso de construcción**
- 2. Verificar documentación**
 - Si la documentación está completa, entonces:
 - Evaluar cumplimiento de normativas
 - Si cumple con las normativas urbanísticas, entonces:
 - Realizar inspección del sitio
 - Si la inspección es satisfactoria, entonces:
 - Aprobar solicitud
 - Enviar notificación de aprobación al solicitante
 - Si la inspección no es satisfactoria, entonces:
 - Rechazar solicitud
 - Enviar notificación de rechazo al solicitante con detalles de las correcciones necesarias
 - Si no cumple con las normativas urbanísticas, entonces:
 - Rechazar solicitud
 - Enviar notificación de rechazo al solicitante con detalles de las normativas no cumplidas
 - Si la documentación no está completa, entonces:
 - Rechazar solicitud
 - Enviar notificación de rechazo al solicitante solicitando la documentación faltante

Fin



2

2

https://governobert.gencat.cat/web/.content/01_Que_es/04_Publicacions/colleccio_govern_obert/GovernObert_9/Govern-Obert-9-cast.pdf

CRITERIOS ESPECÍFICOS PARA LA EVALUACIÓN DEL RIESGO DE LOS ALGORITMOS

Como hemos indicado, los grandes campos que afectan al funcionamiento de los algoritmos son por una parte la metodología utilizada, y por otro el sistema de extracción de datos y su fiabilidad.

ESCALA GENERAL DE RIESGO DE LOS ALGORITMOS SEGÚN LAS FUNCIONES QUE DESARROLLAN

La escala general de niveles de riesgo de la que partimos, obtenida a partir de muchas fuentes y que ha sido consensuada por diferentes actores, tiene que ver directamente con el tipo de acción que desarrolla el sistema de IA. Dentro de este esquema, y teniendo en cuenta las características de los procedimientos de IA, debe tenerse especial cuidado en el control de ciertas categorías de algoritmos, que se detallan a continuación (inicialmente aquellos sistemas declarados de alto riesgo por el RIA):

- Sistemas de identificación biométrica: Utilizados para reconocer características físicas, como huellas dactilares y reconocimiento facial.
- Sistemas que ayudan a coordinar respuestas a desastres y controlan el paso de personas en las fronteras.
- Algoritmos que apoyan en investigaciones, actuaciones judiciales y decisiones judiciales.
- Sistemas que gestionan el acceso a instituciones educativas y programas formativos.
- Herramientas que evalúan el desempeño laboral y facilitan servicios relacionados con el empleo.
- Dispositivos médicos, automóviles, maquinaria y robótica de asistencia y cuidado personal: Tecnologías usadas en el cuidado de la salud, la conducción autónoma, la operación de maquinaria y la asistencia a personas.



Imagen creada con Dall-E

Estos sistemas representarían aproximadamente un tercio de los sistemas de IA utilizados en el sector público.

- Sistemas de control de aplicación de la ley, de priorización de acciones de inspección fiscal y control del fraude, o sistemas de decisión sobre actuaciones que afecten de manera importante a la integridad física y jurídica y a la salud de individuos particulares.
- Sistemas de adjudicación de contratos y subvenciones basándose en criterios de eficiencia y necesidad.
- Sistemas para la gestión de servicios que se presten a los ciudadanos en salud, vivienda, ayudas sociales..., por ejemplo, basándose en la situación económica y familiar de los solicitantes.
- Sistemas para la obtención de información para la elaboración de políticas y toma de decisiones. Algoritmos que analizan datos demográficos y económicos para planificar nuevas políticas de empleo.
- Sistemas para la gestión interna de procesos y recursos de la administración. Por ejemplo, algoritmos que optimizan la asignación de recursos humanos dentro de una entidad pública.
- Sistemas de información e interacción con los ciudadanos para la prestación de servicios, como Chatbots que proporcionan información sobre trámites administrativos y servicios públicos.

EVALUACIÓN DE ALGORITMOS SEGÚN LA TRANSPARENCIA EN SU FUNCIONAMIENTO Y EL NIVEL DE RIESGO DE PRODUCIR DATOS ERRÓNEOS NO DETECTABLES.

La transparencia debe evaluarse caso por caso. Sin embargo, se pueden proponer algunos criterios para evaluar el riesgo según la tecnología utilizada en el algoritmo y su funcionamiento.

Podemos clasificar las técnicas de IA en función del riesgo asociado a su auditabilidad y transparencia. La clave es la posibilidad de que los usuarios puedan contrastar los resultados del sistema de IA mediante observaciones y cálculos simples. Las técnicas más transparentes deberían facilitar la detección de errores evidentes, lo que marca el riesgo de su uso.



Imagen creada con runwayml

A continuación, se presenta una clasificación en *orden descendente* del nivel de riesgo:

1. **Algoritmos opacos o cerrados de contenido no declarado o desconocido. Riesgo alto.** Algoritmos cuyo contenido y funcionamiento no son conocidos ni declarados. Su falta de transparencia dificulta la detección de errores y sesgos, lo que puede generar consecuencias negativas significativas. Por ejemplo: sistemas propietarios donde no se revela cómo se procesan los datos ni los criterios de decisión.
2. **Técnicas de IA generativa (como ChatGPT). Riesgo alto.** Algoritmos que generan contenido a partir de datos no controlables. Su capacidad para crear respuestas basadas en datos variados y no siempre verificables aumenta el riesgo de producir resultados erróneos o sesgados.
3. **Redes neuronales y sistemas de IA basados en aprendizaje profundo sin sistemas de información transparentes,** como redes neuronales que realizan diagnósticos médicos sin explicar cómo llegaron a esas conclusiones. Estos sistemas, a menos que tengan implementados mecanismos de explicabilidad y supervisión, pueden ser difíciles de auditar y entender. Sin embargo, pueden proporcionar información valiosa si se gestionan adecuadamente. Riesgo medio-alto.
4. **Técnicas de agrupamiento, tratamiento de imágenes y algoritmos de aprendizaje no supervisado,** como algoritmos que agrupan imágenes de seguridad sin permitir la verificación directa de su funcionamiento. Riesgo medio.
5. **Algoritmos de regresión, ajuste e inferencia estadística con parámetros de control disponibles.** Como modelos de predicción económica que proporcionan

errores estimados y niveles de confianza, permitiendo la verificación de resultados. Estos algoritmos agrupan datos sin intervención humana, lo que puede dificultar la comprensión de su funcionamiento. La falta de supervisión puede llevar a decisiones no transparentes y menos auditables. Riesgo bajo-medio

6. **Estadísticas descriptivas directas u otros tratamientos matemáticos sencillos.** Técnicas basadas en datos abiertos y fácilmente verificables por cualquier usuario. Estos métodos son transparentes y permiten la detección inmediata de errores masivos. Como los análisis de datos abiertos sobre la contaminación atmosférica, donde los ciudadanos pueden verificar los datos por sí mismos. Riesgo bajo

EVALUACIÓN DE RIESGO DE USO DE ALGORITMOS SEGÚN EL IMPACTO SOCIAL

El impacto de los algoritmos en los ciudadanos no se mide solo por el número de personas afectadas, sino también por la gravedad de las consecuencias de una respuesta errónea. Un error en la programación de un evento no tiene el mismo peso que excluir a alguien de una lista de espera para un tratamiento médico vital. Por lo tanto, la evaluación del impacto debe considerar ambos aspectos. A continuación, se presenta una clasificación de técnicas de IA ordenadas por su nivel de riesgo descendente:

Sistemas autónomos sin supervisión humana: un sistema de control automático de tráfico sin intervención humana. Riesgo: Alto, debido a la falta de intervención humana que puede corregir errores en tiempo real.

1. **Sistemas basados en IA generativa, como ChatGPT** generando contenido basado en datos no controlados. Riesgo: Alto, debido a la capacidad de crear información no verificada y potencialmente errónea.
2. **Redes neuronales no explicables y aprendizaje profundo:** agrupamiento y clasificación dinámica según la adquisición de datos, combinados o no con aprendizaje reforzado. Por ejemplo, Redes neuronales usadas para clasificar perfiles de ciudadanos sin transparencia en los procesos. Riesgo: Alto, por la complejidad y falta de transparencia en sus procesos.
3. **Sistemas de decisión no supervisada,** como análisis dinámico masivo de datos estadísticos a partir de perfiles de ciudadanos u otros datos protegidos. Ejemplo, análisis de datos estadísticos para decidir políticas públicas sin revisión humana. Riesgo: Medio, ya que, aunque manejan grandes volúmenes de datos, la falta de supervisión directa aumenta la posibilidad de errores significativos.
4. **Procedimientos de ajuste de datos mediante modelos lineales.** Por ejemplo, modelos con control de outliers (que son datos que se encuentran significativamente alejados de los demás valores en un conjunto de datos y que

puede indicar una variación inusual o un error) y análisis de errores supervisables. Riesgo: Bajo, debido a la posibilidad de supervisión y corrección de errores.

5. **Estadística descriptiva y operaciones matemáticas elementales.** Así, serían técnicas cuyos códigos se pueden verificar fácilmente y cuyas respuestas frente a entradas particulares son calculables por los usuarios. Por ejemplo, cálculos de estadísticas de tráfico usando datos Riesgo: Muy bajo, por su simplicidad y transparencia.

Esta clasificación ayuda a identificar y gestionar los riesgos asociados con el uso de diferentes técnicas de IA, garantizando así la protección y bienestar de los ciudadanos.



Imagen creada con Dall-E

¿CÓMO EVALUAR EL RIESGO DEL SISTEMA SEGÚN LA CALIDAD DE LOS DATOS?

Evaluar el riesgo según la calidad y el origen de los datos es crucial para garantizar que los sistemas de IA produzcan resultados precisos y confiables. Los datos son el *alimento* de la IA, su calidad es esencial y debe ser controlada cuidadosamente. Ya la AEPD o el Reglamento de IA de la UE exige la calidad de los datos no como una buena práctica, sino como una obligación exigible. Los datos deben ser de alta calidad. Esto implica crear y seguir un protocolo específico para la adquisición de datos. Afortunadamente, ya existen herramientas y estándares para garantizar la calidad de los datos.

Para asegurar la calidad y transparencia de los datos utilizados en los algoritmos públicos, se deben seguir estas pautas:

- *Formato y procesabilidad:* Los datos deben estar en un formato abierto y fácil de procesar. Usar una codificación de caracteres reconocida internacionalmente, como UTF-8.
- *Estandarización y organización:* Las columnas deben tener nombres descriptivos y estandarizados en minúsculas (ASCII). Especificar claramente los valores nulos.
- *Recogida y almacenamiento:* Implementar un estándar de recolección y almacenamiento de datos que permita detectar duplicados fácilmente.
- Utilizar un vocabulario de referencia estandarizado.
- *Cantidad y contexto y geográficos:* Asegurar una cantidad suficiente de datos para proporcionar contexto adecuado. Usar el estándar ISO 8601 para codificar fechas.
- Utilizar el punto como separador decimal.
- Mantener una escala numérica constante a lo largo del tiempo.
- Proporcionar el mayor nivel posible de detalle en los datos.
- Incluir las coordenadas geográficas (latitud y longitud) en columnas separadas.

Las propiedades de calidad de los datos que deben observarse, según la norma ISO/IEC 25012, son:

- *Exactitud:* Los datos deben reflejar correctamente la realidad.
- *Completitud:* No deben faltar datos necesarios.
- *Consistencia:* Los datos deben ser coherentes entre diferentes sistemas y momentos.
- *Credibilidad:* La fuente de los datos debe ser confiable.
- *Actualidad:* Los datos deben estar actualizados.
- *Accesibilidad:* Los datos deben ser fáciles de obtener y utilizar.
- *Conformidad:* Los datos deben cumplir con las normas y estándares establecidos.
- *Confidencialidad:* Se debe proteger la privacidad de los datos.
- *Eficiencia:* Los datos deben ser manejables sin consumir excesivos recursos.
- *Precisión:* Los datos deben ser específicos y detallados.
- *Trazabilidad:* Debe ser posible seguir el origen y los cambios de los datos.
- *Comprensibilidad:* Los datos deben ser claros y fáciles de interpretar.



Un aspecto importante, pero complejo de controlar, es asegurar que los datos estén libres de sesgos (por ejemplo, por edad, sexo, grupo social), lo que podría producir soluciones erróneas o discriminatorias. Esto no es solo un problema técnico, sino también humano. La única manera de garantizar la adecuación de los datos es mediante supervisión humana. La alimentación de los sistemas algorítmicos debe estar supervisada por un organismo que gestione su validez, basándose en la evaluación de analistas expertos que revisen cada caso individualmente.

Este control debe implementarse en paralelo con la supervisión de los algoritmos, integrando la calidad de los datos con el nivel de riesgo de los algoritmos utilizados para evaluar el riesgo global de una metodología. No es lo mismo extraer conclusiones de un sistema de IA con datos bien recopilados y formateados (por ejemplo, CSVs sin datos faltantes y bien formateados) que con datos incompletos y en formatos complicados (por ejemplo, PDFs con texto intercalado).

Las soluciones aportadas por los sistemas de IA deben analizarse desde el principio, asegurando el uso de datos de calidad y sin sesgos para su funcionamiento. Se deben considerar sistemas integrales basados en IA, que combinen la adquisición de datos con las técnicas algorítmicas utilizadas:

Sistemas integrales basados en IA = Sistema de adquisición de datos + Técnicas de IA utilizadas.

¿CÓMO VALORAR DEL UNO AL CINCO EL RIESGO GLOBAL DE LOS SISTEMAS INTEGRALES DE IA EN EL SECTOR PÚBLICO?

Como consecuencia de los diferentes aspectos comentados, la estimación del riesgo debe hacerse teniendo en cuenta todos los factores indicados. Este proceso debe considerar varios factores, agrupándolos en categorías con una puntuación del 0 al 5, donde 0 es sin riesgo y 5 es riesgo máximo. Las categorías a considerar son:

- *Calidad de los datos*: Verifica la exactitud y completitud de los datos.
- *Veracidad y sesgo*: Evalúa posibles sesgos en los datos.
- *Tecnología utilizada*: Analiza la fiabilidad, complejidad, opacidad y transparencia de la tecnología.
- *Impacto en la ciudadanía*: Considera las consecuencias de errores en la población e impacto en sus derechos

El comité de expertos debe desarrollar una metodología estandarizada que se pueda mejorar con el tiempo. Un ejemplo práctico sería la asignación de ayudas a la vivienda en el que se podrían establecer ciertos umbrales:

- Calidad de los datos: ≤ 3
- Veracidad y sesgo: ≤ 3
- Tecnología utilizada: ≤ 4
- Impacto en la población: ≤ 4

Para sistemas más sensibles, como el control de fraudes financieros, los requisitos pueden ser más estrictos:

- Calidad de los datos: ≤ 2
- Veracidad y sesgo: ≤ 1
- Tecnología utilizada: ≤ 1
- Impacto en la población: ≤ 3

Este enfoque asegura que todos los aspectos críticos se evalúen y gestionen adecuadamente, minimizando riesgos y optimizando la confiabilidad del sistema.

IV. ¿QUÉ INFORMACIÓN DEBE APORTARSE PARA EL REGISTRO DE ALGORITMOS DE IA?

La cantidad y el detalle de la información a proporcionar sobre los algoritmos públicos dependerán del nivel de riesgo asignado a cada sistema.³ La normativa valenciana es bastante limitada en este aspecto, comparada con las normativas y mejores prácticas de países como Reino Unido, Francia o Países Bajos. Esta normativa solo indica que la información debe incluir:

- Una descripción comprensible del diseño y funcionamiento del algoritmo.
- El nivel de riesgo asociado.
- Un punto de contacto para consultas, en línea con los principios de transparencia y explicabilidad.

Será responsabilidad del órgano técnico o comité de registro público de algoritmos determinar los elementos de información que se deben proporcionar, según el nivel

En muchas propuestas actuales para el registro de algoritmos se incluye una ficha técnica que deben completar los responsables de implementar cada nuevo sistema de IA. Esta ficha debe cubrir todas las características relevantes, desde la calidad de los datos de entrada hasta las propiedades del algoritmo y la evaluación de su riesgo e impacto. La finalidad es proporcionar una evaluación integral del riesgo que permita establecer las medidas de control adecuadas.

- Los apartados que se deben incluir en la ficha son los siguientes:
- Calidad esperada de los datos de entrada, medidos según el sistema explicado anteriormente u otro similar.
- Evaluación por expertos del riesgo de falsedad o sesgo de los datos de entrada.
- Tipo de IA utilizada.
- Estimación inicial, según la lista indicada en la Sección 2, del riesgo de uso y del impacto.

³ Sobre la información a facilitar, un estudio exhaustivo, Cotino Hueso, L. "Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida", *Revista Española de la Transparencia*, Núm. 16 (Primer semestre. Enero - junio 2023). [acceso](#)

PROPUESTA DE FICHA DE REGISTRO

Basándonos en el análisis de los registros de otras administraciones, proponemos la siguiente ficha. Esta es más detallada que la utilizada en la AOC (Administración abierta de Cataluña)⁴ pero menos exhaustiva que la del Reino Unido, buscando un equilibrio que no *desanime* a los encargados de su cumplimentación.

Ficha de Registro de Algoritmos Transparentes. Propuesta.

<i>Información General</i>	
Nombre del Algoritmo	Nombre público y/o código interno
Descripción y Contexto	Describir el contexto de desarrollo, la problemática abordada, la solución implementada y el estado actual del algoritmo
Objetivos del Algoritmo	Detallar los objetivos específicos y las funciones para las que fue desarrollado el algoritmo
Estatus de Desarrollo	Incluir estados como en definición, planificado, en desarrollo, piloto, implementado, no operativo.
<i>Desarrollo e Implementación</i>	
Administración Responsable	Entidad o departamento que gestiona el uso del algoritmo.
Entidad Desarrolladora	Nombre y datos de la empresa o personas que desarrollaron el algoritmo
Fechas Clave	<ul style="list-style-type: none"> • Fecha de inicio de uso • Fecha de prueba piloto • Estimaciones de futuros desarrollos o finalización
Tipo de Algoritmo	Describir la tecnología específica utilizada (p. ej., aprendizaje automático supervisado, procesamiento de lenguaje natural)
<i>Datos y Metodología</i>	

⁴ <https://www.aoc.cat/es/proyecto-innovacio/transparencia-en-lus-dalgorismes-dintelligencia-artificial-a-laoc/>

Datos Utilizados	<ul style="list-style-type: none"> • Detalles sobre los conjuntos de datos, incluyendo la fuente, la relevancia y el tratamiento de datos personales. • Procesamiento de Datos: Explicar cómo se procesan los datos dentro del algoritmo, incluyendo pasos clave y arquitectura del sistema
Beneficios	Beneficios: Describir los beneficios específicos tanto para la administración como para los ciudadanos.
Riesgos	Riesgos Potenciales: Identificar riesgos como discriminación, amenazas a la privacidad, manipulación, etc
Supervisión	Supervisión Humana: Detallar el nivel de intervención humana en las decisiones del algoritmo
Transparencia	<ul style="list-style-type: none"> • Transparencia y Accesibilidad: Información sobre cómo los ciudadanos pueden acceder a datos sobre el algoritmo, incluyendo un enlace a un portal de transparencia si está disponible. • Visibilidad y Conciencia Pública: Explicar cómo se informa a los ciudadanos sobre el uso del algoritmo
Cumplimiento Normativo y Auditoría	Normativas Aplicables: Confirmar el cumplimiento de leyes de protección de datos y otras regulaciones relevantes
Auditorías y Evaluaciones de Impacto	Detalles de las auditorías realizadas, incluyendo la frecuencia, los hallazgos y las métricas de rendimiento.
Participación y Retroalimentación	<ul style="list-style-type: none"> • Mecanismos de Retroalimentación: Procesos para que los ciudadanos y usuarios afectados puedan proporcionar comentarios o disputar decisiones. • Impacto en la Política de Igualdad: Evaluación del impacto del algoritmo en la igualdad y no discriminación, especialmente en contextos laborales.
Documentación y Recursos Adicionales	<ul style="list-style-type: none"> • Información de Contacto: Detalles de contacto para consultas adicionales. • Enlaces y Recursos: Bibliografía y enlaces a documentos de política, estudios de caso y otros materiales relevantes.

V. UN “PASEO” POR REGISTROS DE ALGORITMOS PÚBLICOS Y MEJORES PRÁCTICAS

ALGUNOS ESTÁNDARES Y REGISTROS DESTACADOS

ALGORITHMIC TRANSPARENCY STANDARD DE EUROCITIES

Este estándar de Eurocities⁵, que es la red de más de 200 ciudades en 38 países, que representan 130 millones de personas. En principio participan en la iniciativa Amsterdam, Barcelona, Brussels, Eindhoven, Mannheim, Rotterdam y Sofia.

Es un conjunto de categorías de información compartidas que las ciudades que permite a los ciudadanos comparar diferentes algoritmos dentro y entre ciudades, apoyando así la documentación de decisiones y suposiciones tanto para la gestión de la gobernanza de la IA como para proporcionar una transparencia significativa de manera estandarizada. El estándar está diseñado para documentar las decisiones y suposiciones relacionadas con la gestión y gobernanza de la IA, proporcionando una transparencia significativa y estandarizada.

ICO DE REINO UNIDO: ALGORITHMIC TRANSPARENCY RECORDING STANDARD

El Algorithmic Transparency Recording Standard - Guidance for Public Sector Bodies del ICO de Reino Unido⁶

Está dirigido a organizaciones del sector público que usan herramientas algorítmicas con un impacto significativo en la toma de decisiones o que interactúan directamente con el público. Es obligatorio para los departamentos del gobierno central y se recomienda su uso en el sector público en general. La guía está estructurada en tres pasos: preparación, redacción y publicación del registro de transparencia algorítmica. Estos pasos incluyen la identificación de herramientas relevantes, la recolección y redacción de la información necesaria, y el proceso para subir y actualizar los registros en el repositorio público de GOV.UK.

Incluye un Excel de la máxima exhaustividad

⁵ <https://www.algorithmregister.org/standard>

⁶ Su última versión es de 5 de enero 2023.

Algorithmic Transparency Recording Standard – Template
 ([https://assets.publishing.service.gov.uk/media/664ba9d9f34f9b5a56adcafc/Algorithmic Transparency Recording Standard - Template.xlsx.xlsx](https://assets.publishing.service.gov.uk/media/664ba9d9f34f9b5a56adcafc/Algorithmic%20Transparency%20Recording%20Standard%20-%20Template.xlsx.xlsx))

GOV.UK

Register to vote Register by 18 June to vote in the General Election on 4 July.

Home > Guidance for organisations using the Algorithmic Transparency Recording Standard

Central Digital & Data Office | Department for Science, Innovation & Technology

Guidance
Algorithmic Transparency Recording Standard - Guidance for Public Sector Bodies
 Published 5 January 2023

Contents

1. Overview
2. Using this guidance
3. Step 1: Prepare to complete an algorithmic

1. Overview

What is the Algorithmic Transparency Recording Standard?

The Algorithmic Transparency Recording Standard (ATRS) is a framework for capturing information about algorithmic tools

Documents



[Algorithmic Transparency Recording Standard - Template](#)

MS Excel Spreadsheet, 333 KB

This file may not be suitable for users of assistive technology.

▶ [Request an accessible format.](#)



[Algorithmic Transparency Recording Standard - Template \(Google Sheets\)](#)

<https://docs.google.com/spreadsheets/d/1mZbQoT4RkE84RMx-ENTcPM22DOWLcQy-Vwga18Bieyk/edit#gid=239944889>

A	B	C	D	F	G	H
Algorithmic Transparency Recording Standard v3.0						
Section 2.1. Owner and Responsibility						
In this section, you should provide information about accountability for the development and deployment of the tool.						
Section complete						
#	Field	Prompt	Answer	Complete?		
2.1.1	Organisation or department	Provide the full name of the organisation, department or public sector body that carries responsibility for use of the algorithmic tool.	Enter organisation here	n		
2.1.2	Team	Provide the full name of the team that carries responsibility for use of the algorithmic tool.	Enter names here	n		
2.1.3	Senior responsible owner	Provide the role title of the senior responsible owner for the algorithmic tool.	Enter name here	n		
2.1.4	External supplier involvement	Have external suppliers been involved in the development and operation of the tool?	Select from options	n		
2.1.4.1	External supplier *	Provide the name of any external organisation that has been contracted to develop the whole or parts of the algorithmic tool.	Enter answer here. Type "N/A" if no external supplier.	n		
2.1.4.2	Companies House Number *	If available, provide the Companies House number of the external organisation that has	Enter answer here. Type "N/A" if no external supplier.	n		

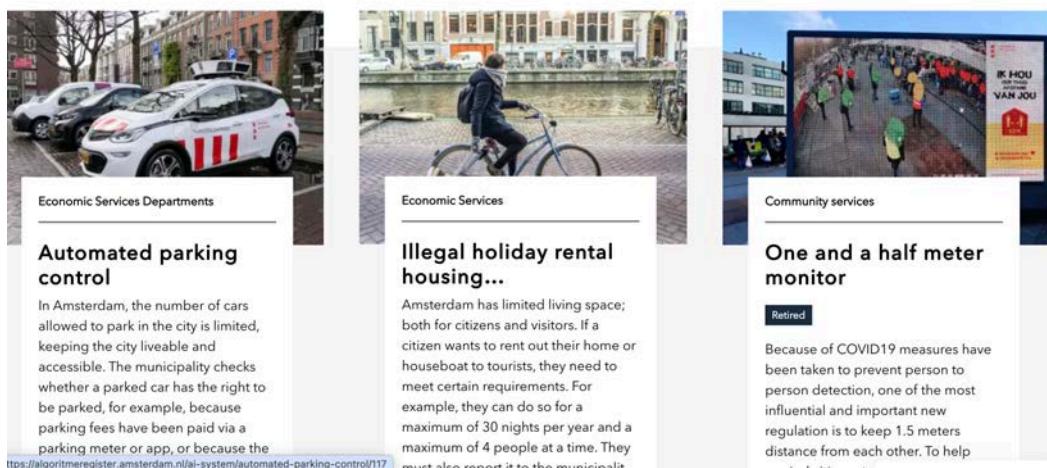
LOS REGISTROS DE ALGORITMOS DE ÁMSTERDAM- HELSINKI

A este respecto, cabe seguir el documento, Meeri Haataja, Linda van de Fliert and Pasi Rautio, *Whitepaper, Public AI Registers, Realising AI transparency and civic participation in government use of AI*, 2020. <https://ai.hel.fi/wp-content/uploads/White-Paper.pdf>

El Registro de Algoritmos de Ámsterdam permite a cualquier persona consultar y entender los algoritmos empleados por la ciudad. El registro se actualiza con nuevos algoritmos y sistemas de IA a medida que son implementados. Cada entrada incluye una descripción detallada del algoritmo, explicando su propósito y funcionamiento específico dentro de la administración municipal. Se explica el uso específico y detalles técnicos relevantes sobre los algoritmos, como el tipo de datos que procesan y los métodos utilizados para entrenar los modelos de IA. Se centra en el impacto potencial de estos algoritmos en los ciudadanos.

Dicho lo anterior, lo cierto es que sólo incluye información sobre algoritmos utilizados en varios servicios municipales: control de estacionamiento automatizado,⁷ alquileres ilegales de viviendas vacacionales,⁸ y ya retirado el Sistema One and a Half Meter Monitor en la época del Covid⁹.

Learn about the use cases where we currently utilise algorithmic systems as part of our city services.



El *Helsinki Algorithm Register* <https://ai.hel.fi/en/ai-register/> tiene exactamente la misma apariencia visual que el de Ámsterdam. El mismo cuenta con más casos, si bien

⁷ El control automatizado de estacionamiento en Ámsterdam usa autos con cámaras y software de reconocimiento de objetos para escanear placas en más de 150,000 espacios. El proceso incluye: escaneo: Autos escanean placas de autos estacionados. Verificación: Placas verificadas contra el Registro Nacional de Estacionamiento. Inspección: Casos no pagados se envían a inspectores humanos para revisión. Si no se encuentra una razón válida para no pagar, se emite una multa.

⁸ Illegal Holiday Rental Housing, desde julio de 2020 se utiliza un algoritmo que prioriza los informes de posibles alquileres ilegales, ayudando a los empleados del departamento de Vigilancia y Control en sus investigaciones.

⁹ Sistema para concienciar a los ciudadanos sobre el distanciamiento social. Utilizaba cámaras y pantallas para mostrar una imagen con una superposición de un emoticón según la distancia mantenida: verde para más de 1.8 metros, naranja para 1.5-1.8 metros y rojo para menos de 1.5 metros. Este sistema no almacenaba información visual y se utilizó temporalmente durante la ordenanza de emergencia por COVID-19.

son todo chagbots: Parking Chatbot, Outdoors Chatbot Urho, City Executive Office, International House Helsinki's Chatbot , Sotebotti Hester chatbot, Oodi's Book Recommendation Chatbot, Rental Apartment Search Chatbot y un canal de contacto electrónico para los servicios de gestión financiera de Helsinki. No muy relevantes. Ahora bien, de cada algoritmo se ofrece información de Datasets,

Data processing, medidas de no discriminación, Human oversight y Risk management: Estrategias y procedimientos establecidos para identificar, evaluar y mitigar los riesgos.

Helsinki City of Helsinki AI Register

AI Register Get to know AI Register Participate in a survey

What is AI Register?

AI Register is a window into the artificial intelligence systems used by the City of Helsinki. Through the register, you can get acquainted with the quick overviews of the city's artificial intelligence systems or examine their more detailed information based on your own interests. You can also give feedback and thus participate in building human-centered AI in Helsinki.

[Get to know AI Register](#)

- Parking chatbot**
Housing and environment
The parking chatbot is a customer service channel of city's parking services. Service provides automated answers to the parking-related questions of city residents and visitors. The service is available at the city parking website of Helsinki. The service aims to improve the availability and the user experience of...
> [Read more](#)
- Outdoors chatbot Urho**
Outdoors chatbot Urho is a 24-hour customer service channel of the Helsinki City Information aimed at improving the accessibility of customer service and the customer experience and increasing the interactivity of the self-service. The service provides relevant information to each customer's specific questions faster than by...
> [Read more](#)
- City Executive office**
International House Helsinki...
International House Helsinki's chatbot Into is a 24/7 customer service channel, offering a wide range of information on the official services offered by IHH and advice to support the settlement of those who have moved to the capital region from abroad. With the help of the service, customers have faster access to International House Helsinki's wide...
> [Read more](#)
- Sotebotti Hester**
Social services, Health Care and Rescue Services Division
Sotebotti Hester is a chatbot for social services, health care and rescue services division. Hester contains different knowledge data bases that are combined into one chatbot. The same answer can be used in many different
- Intelligent material...**
Culture and leisure
IMMS (Intelligent Material Management System) is an intelligent material management system for the entire library collection. The City Library's collection contains approximately 1.8 million items. An intelligent material
- Oodi's book recommendation...**
Culture and leisure
Oodi is Central Library Oodi's recommendation chatbot. The service recommends books from Oodi's collection according to the customer's interest and feedback. The service is

REGISTRO DE ALGORITMOS DE PAÍSES BAJOS

El Registro de Algoritmos de los Países Bajos ya cuenta con 385 algoritmos (junio 2024). Permite de hecho, descargarse cada información o la de todos en un archivo conjunto de xls (<https://algoritmes.overheid.nl/en>) recopila y publica los algoritmos utilizados por diferentes organismos gubernamentales, tiene como objetivo proporcionar a la ciudadanía una comprensión detallada del funcionamiento de estos algoritmos y su impacto en la toma de decisiones gubernamentales. El registro se centra en aquellos algoritmos que tienen un impacto significativo en la población. Las organizaciones gubernamentales añaden continuamente información sobre nuevos algoritmos, en un esfuerzo colaborativo. Se anima a más entidades a unirse a esta iniciativa para seguir expandiendo y mejorando el registro.



The screenshot shows the homepage of the Dutch Algorithm Register. At the top, there is a navigation bar with 'Home', 'Algoritmes', and 'Organisaties' tabs, and a language selector set to 'Taal: Nederlands'. The main heading is 'Het Algoritmeregister van de Nederlandse overheid'. Below this is a search bar with the text 'Vind een van de 385 algoritmes' and a search button labeled 'Zoeken'. A sub-section titled 'More detailed information on the system' contains a paragraph and a list of categories: 'Datasets', 'Data processing', 'Non-discrimination', 'Human oversight', and 'Risk management', each with a 'Show More' button and a dropdown arrow. A small feedback icon is visible at the bottom right of the list.

Was this information useful?

GUÍA ETALAB FRANCIA

La guía Etalab del Gobierno *Expliquer les algorithmes publics*
<https://guides.etalab.gouv.fr/pdf/guide-algorithmes.pdf> y
<https://etalab.github.io/algorithmes-publics/guide.html>

La guía del Etalab (2022) titulada "Expliquer les algorithmes publics" es un documento creado para acompañar a las administraciones en el uso responsable y transparente de los algoritmos públicos. La guía da un marco claro y accesible para explicar cómo funcionan los algoritmos utilizados por las administraciones públicas, y por lo que más interesa, para garantizar la transparencia y la responsabilidad en su implementación.

algorithmes-publics



Guide des algorithmes publics

Cabe recordar que Código de Relaciones entre el Público y la Administración (CRPA) francesa concreta las administraciones y los tratamientos sujetos a la transparencia algorítmica: administraciones estatales, colectividades, organismos públicos o privados con misión de servicio público, que utilizan tratamientos algorítmicos para tomar decisiones administrativas individuales sobre personas físicas o morales. *Respecto de los tratamientos automatizados*, desde el 1 de julio de 2020, se han de informar expresamente bajo pena de nulidad de la actuación administrativa. El uso de algoritmo implica *obligaciones de transparencia sobre información general*: Publicar en línea las reglas que definen los principales tratamientos usados en decisiones individuales (aplicable a administraciones con más de 50 empleados o colectividades con más de 3500 habitantes). También incluir en documentos y online la finalidad del tratamiento, derecho de comunicación y cómo ejercerlo. Ya a solicitud del interesado, hay que proporcionar detalles sobre el grado y modo de contribución del algoritmo en la decisión, datos tratados, fuentes, parámetros y operaciones realizadas.

Algorithmes publics aujourd’hui publiés

- En cas d’absence de licence, le droit s’applique normalement et le code source est librement réutilisable par tous.

Thème / objet	Producteur	URL de publication	Informations complémentaires	Licence
Calcul des Allocations Familiales	CNAF	Nextimpact	Nextimpact	N/A*
Calcul de la taxe d’habitation	DGFIP	Github	Blog Etalab	CeCILL v2.1
Parcours Sup	MESRI	Framagit	README	GNU GPL v3
Impôt sur le revenu	DGFIP	Gitlab Adullact	README	CeCILL v2.1
Taxe foncière	DGFIP	Gitlab Adullact	README	CeCILL v2.1

OBSERVATORIO IA EN SALUD (CATALUÑA)

Observatorio de la IA en el ámbito de la salud ciertamente tiene otra naturaleza.

<https://iasalut.cat/es/observatori/>

<https://iasalut.cat/es/observatori-ia-en-salut/observatori-dintelligencia-artificial-en-salut/>

Contiene en teoría un repositorio de herramientas de IA para identificar todas las herramientas que se están desarrollando o empleando en el ecosistema catalán y analizar su grado de implantación y desarrollo, así como el grado de escalabilidad a nivel sistémico. En junio de 2024 sólo se puede acceder a la versión en catalán (<https://iasalut.cat/observatori-ia-en-salut/eines-ia-salut/>) y que contiene un “Formulario de registro de herramientas de IA”.

En el mismo se registran los cambios de datos de contacto y del equipo, datos generales de la herramienta, datos específicos de la herramienta, nivel de madurez de la herramienta, datos técnicos de la herramienta y la información adicional.

Eines d'IA en salut

Inici / Observatori IA en salut

Inici **Presentació Observatori** **Eines d'IA en salut** **Documents de referència**



Per identificar i analitzar les eines d'IA en salut que s'estan emprant o desenvolupant en l'ecosistema català, s'ha creat un formulari dirigit a les entitats del Sistema de Salut de Catalunya, centres de recerca, universitats i empreses involucrades en el disseny, desenvolupament, implantació o distribució de les eines d'IA.

Formulario de herramientas de Inteligencia Artificial en Salud - Observatorio de IA en Salud



Información básica del Formulario de herramientas de IA en Salud *

El **Programa Salut/IA**, aprobado en la resolución SLT/954/2023 del pasado 19 de marzo de 2023, tiene como finalidad la creación de un entorno facilitador para la innovación en el ámbito de la salud mediante el desarrollo y la implementación de soluciones de inteligencia artificial (IA) para la mejora de la salud de la ciudadanía. Para hacerlo posible, el programa ha desplegado una serie de líneas estratégicas en torno a tres ejes de actuación: la innovación, la evaluación y la implantación. Si desea consultar más información sobre el Programa, puede acceder a su web: <https://iasalut.cat/>

El Programa impulsa el **Observatorio de IA en Salud**, un espacio de referencia para la innovación en IA y por la medida de su grado de implantación. Si desea más información sobre el Observatorio, puede acceder al siguiente enlace: <https://iasalut.cat/observatori-ia-en-salut/>

El objetivo de este formulario es identificar y analizar las iniciativas de IA que se están desarrollando o empleando en el ámbito de la salud en Cataluña.

Este formulario tiene una duración aproximada de unos 30 minutos. Puede guardar los cambios y reanudarlos más tarde pulsando el botón "Guardar y continuar más tarde" al final de cada página. Una vez completado y enviado el formulario, dispondrá de 10 días para editar las respuestas.

He leído detenidamente la información anterior.

UNA VALORACIÓN DE ESTOS ESTÁNDARES Y EXPERIENCIAS

INFORMACIÓN BÁSICA COMÚN

En resumen, la mayoría de los registros incluyen información básica sobre el algoritmo, la organización responsable, y datos de contacto. Específicamente, todos incluyen datos sobre la propiedad y responsabilidad del algoritmo, la descripción del funcionamiento y tipo de IA utilizada, así como la evaluación del impacto y riesgos asociados.

En concreto, es común incluir la siguiente información:

- *Nombre del algoritmo*: Identificación específica del algoritmo.
 - *Descripción corta*: Resumen breve del propósito y funcionamiento del algoritmo.
 - *Organización responsable*: Entidad que desarrolló o implementa el algoritmo.
 - *Contacto (e-mail o teléfono)*: Información de contacto para consultas o más detalles.
- *Propiedad y Responsabilidad*
 - *Persona u órgano responsable dentro de la administración*: Identificación de quién o cuál departamento es responsable del algoritmo.
 - *Proveedores externos y su nivel de intervención*: Información sobre terceros involucrados y su grado de participación.
- *Datos del Algoritmo*
 - *Descripción de cómo funciona el algoritmo*: Explicación del mecanismo y lógica del algoritmo.
 - *Tipo de IA utilizada*: Clasificación del tipo de inteligencia artificial aplicada (por ejemplo, aprendizaje supervisado, redes neuronales).
 - *Objetivo principal*: Finalidad principal para la cual se utiliza el algoritmo.
- *Datos de Entrada y Entrenamiento*
 - *Descripción de los datos de entrada*: Detalle de los tipos de datos que alimentan el algoritmo.
 - *Fuente de los datos*: Origen de los datos utilizados, como bases de datos internas, datos públicos, etc.
 - *Métodos de entrenamiento del algoritmo*: Técnicas y procedimientos utilizados para entrenar el algoritmo.
- *Evaluación de Impacto y Riesgos*
 - *Evaluaciones de impacto en derechos humanos (IAMA)*: Análisis de cómo el uso del algoritmo podría afectar los derechos humanos.

- *Evaluaciones de impacto en la protección de datos (DPIA)*: Examen de cómo se maneja la privacidad y protección de datos personales.
- *Riesgos identificados y medidas de mitigación*: Potenciales riesgos asociados con el uso del algoritmo y las estrategias para mitigarlos.

INFORMACIÓN DISTINTIVA Y LOS ESTÁNDARES MÁS CONCRETOS

Después de esta información común, se puede observar que algunos estándares o registros incluyen información distintiva. La información específica para cada registro-web es diversa, destacándose en cada uno detalles únicos como la fase de desarrollo (Observatorio IA en Salud), la frecuencia de uso (Países Bajos), y el contexto global (Guía Etalab Francia). En general, los registros de ICO Reino Unido y la Guía Etalab Francia son los más exigentes y contienen la información más específica. Ambos proporcionan una cobertura completa de los aspectos técnicos, éticos y de impacto de los algoritmos, asegurando que se cumplan altos estándares de transparencia y responsabilidad en el uso de IA en el sector público.

En esta dirección, el *Algorithmic Transparency Standard (ATS)* destaca por su nivel de detalle y exigencia en cuanto a la documentación de decisiones y supuestos para la gestión de la gobernanza de la IA. Incluye una descripción completa del funcionamiento del algoritmo, los datos de entrada y entrenamiento, y una evaluación exhaustiva de impacto y riesgos. Incluye información específica sobre Área geográfica: región o áreas donde se aplica el algoritmo; proceso de toma de decisiones: descripción de cómo el algoritmo influye en las decisiones o alternativas consideradas: otras opciones que fueron evaluadas antes de implementar el algoritmo.

También es muy exigente los estándares del *Information Commissioner's Office del Reino Unido*, el más completo.¹⁰ Requiere detalles sobre la organización responsable, los proveedores externos y su nivel de intervención. Adicionalmente, pone un fuerte énfasis en la evaluación de impacto en derechos humanos y la protección de datos, lo cual añade una capa de seguridad y ética en el uso de algoritmos. Entre otras muchas incluye el nivel de intervención del proveedor externo o detalles sobre cómo los proveedores externos acceden y almacenan los datos o una evaluación de impacto en derechos fundamentales.

La *Guía Etalab de Francia* se distingue por su exhaustividad en la descripción del contexto global y la inclusión de evaluaciones detalladas de impacto.

Este registro no solo cubre los aspectos técnicos y de datos, sino que también aboga por una mayor explicabilidad y transparencia en el diseño y funcionamiento de los

algoritmos. Incluye referencias a evaluaciones de impacto en derechos humanos y de protección de datos.

Los registros de Ámsterdam y Helsinki destacan por su enfoque en el impacto social de los algoritmos. Incluyen detalles sobre el objetivo del algoritmo, su uso en procedimientos administrativos y su impacto en los ciudadanos. ¿Cómo se está utilizando? frecuencia con la que se utiliza Hay que informar de evaluaciones de impacto en derechos humanos, o la forma en la que el algoritmo se integra en el procedimiento y áreas específicas de aplicación del algoritmo. También los Departamentos dentro de la administración pública responsables del algoritmo.

El registro de los *Países Bajos* es conocido por su vinculación a la evaluación del impacto en la privacidad y derechos humanos por su claridad y detalle en la descripción del uso y la frecuencia de los algoritmos. También proporciona información esencial sobre los *desarrolladores* la responsabilidad y la gestión de los algoritmos

El caso del *Observatorio IA en Salud (Cataluña)* es detallado en cuanto a las tecnologías utilizadas y las fases de desarrollo del algoritmo, la especialidad y áreas de la salud donde se utiliza el algoritmo. Evaluación de impacto en la protección de datos y colectivos a los que afecta.

VI. PARA ACABAR ¿CÓMO IMPLEMENTAR UN REGISTRO DE ALGORITMOS PÚBLICOS?

Implementar un registro de algoritmos públicos es un proceso complejo que requiere una gobernanza interna adecuada. A continuación, se detallan algunas ideas y pasos necesarios para lograrlo:

¿Qué es necesario para implementar el registro?

- Establecer una organización dedicada

La creación de una entidad específica, como la "Oficina de Registros de Algoritmos" (ORA), es fundamental. Esta oficina debe estar directamente vinculada a la presidencia de la Comunidad Autónoma o la Alcaldía para asegurar un apoyo transversal y la colaboración de múltiples departamentos y unidades.

- Recopilación y evaluación de información

1. Recopilación de datos:

Auto búsqueda: Revisar registros de actividades de tratamientos de datos, catálogos internos de decisiones automatizadas, normativas, circulares e informes para identificar posibles algoritmos.

Requerimientos de ayuda/información: Enviar correos electrónicos y órdenes directas a través de las distintas consellerías, departamentos y utilizar canales internos para recopilar información sobre sistemas algorítmicos.

2. Clasificación y evaluación de riesgos:

Evaluar y clasificar los algoritmos según el riesgo asociado: Los riesgos pueden escalar de 1 a 5, dividiéndose en alto, medio y bajo. El Consejo del Registro de Algoritmos Públicos (CRAP) es responsable de esta tarea, o cuanto menos de definir la metodología, instrucciones y directrices para la evaluación y supervisión.

3. Incorporación al registro:

Incluir los algoritmos en el registro: Determinar qué información debe ser pública, basándose en la evaluación de riesgos.

4. Actualización continua:

Evaluar y actualizar continuamente el registro: Incluir nuevos algoritmos y ajustar la información existente según cambios en los riesgos o nuevas normativas.

- **Gobernanza y transparencia**

Se deben establecer medidas de gobernanza algorítmica en función del riesgo asignado. La transparencia es clave para la confianza pública, por lo que es esencial definir claramente qué información será accesible y cómo se actualizará regularmente.

- ¿Quién debe estar involucrado?

Oficina de Registros de Algoritmos (ORA). La ORA debe estar compuesta por personal administrativo especializado y, en caso necesario, contar con expertos en transparencia y gobernanza de algoritmos. Sus funciones incluyen:

- Buscar y mapear los algoritmos públicos.
- Asistir al CRAP en la aplicación de medidas.
- Asegurar la correcta gestión del registro.

Consejo del Registro de Algoritmos Públicos (CRAP). Este consejo es responsable de:

- Evaluación de riesgos:
- Clasificar los algoritmos según su riesgo y definir la información que debe publicarse. El Consejo CRAP es responsable de esta tarea o en su caso de la definición de la metodología e instrucciones y directrices para la evaluación
- Resolución de consultas:
- Responder a las consultas de la ORA y otras entidades implicadas.

El CRAP debería tener una composición híbrida, incluyendo:

Expertos académicos: Representantes de universidades en derecho, inteligencia artificial, ciencia de datos y matemáticas.

Sociedad civil: Representantes de organizaciones no gubernamentales y empresas tecnológicas.

Responsables públicos: Presidido por la entidad con competencias en transparencia y el delegado de protección de datos como secretario.

Implementación práctica

- Evaluación y asignación del riesgo

Riesgo alto: Los sistemas de alto riesgo se registrarán con una mayor exigencia de publicidad activa y deberán incluir informes detallados de evaluación de impacto.

Riesgo medio: Requieren un seguimiento continuo y pueden ser reevaluados para determinar si deben ser reclasificados.

Riesgo bajo: No se requiere actuación adicional más allá de su inclusión en el registro.



Imagen creada con Dall-E

- Actualización y seguimiento

Informe anual del CRAP: Propuestas de cambios y mejoras en la transparencia algorítmica.

Revisión de riesgos: Evaluación continua de los sistemas clasificados como de riesgo medio.

ANEXOS

ANEXO I. SISTEMAS DEL AYUNTAMIENTO DE BARCELONA EN SU PRIMERA GUÍA

El Ayuntamiento de Barcelona ha aplicado sistemas de IA en¹¹:

- *Dirección de servicios de información y atención a la ciudadanía un servicio de soporte para la clasificación de incidencias.* IRIS es un servicio implementado en el Ayuntamiento de Barcelona que permite a los ciudadanos comunicar incidencias o reclamaciones al ayuntamiento para su resolución. Los usuarios deben clasificar las incidencias mediante un sistema de categorías proporcionado por la aplicación, siendo crucial esta clasificación para dirigir eficazmente la incidencia al departamento adecuado y acelerar la respuesta. Sin embargo, errores en esta clasificación pueden llevar a respuestas inadecuadas y retrasos. Para mejorar este proceso, se ha desarrollado un módulo llamado MARIO, que utiliza algoritmos de aprendizaje automático y procesamiento de lenguaje natural para ayudar en la clasificación. MARIO, que actualmente está en fase de prueba, ha logrado reducir significativamente los errores de clasificación, aumentando la precisión desde un 50% de necesidad de reasignación manual a un éxito del 85%. Este algoritmo se ha desarrollado con herramientas de código abierto como Python, Scikit-learn y Pandas, y se ha entrenado con datos previamente validados.
- *Servicios Sociales. Área responsable: Derechos Sociales, Justicia Global, Feminismos y LGTBI.* El Área de Derechos Sociales del Ayuntamiento de Barcelona gestiona una media de cincuenta mil primeras visitas anuales, atendiendo a personas con problemas económicos, de dependencia, mentales, de alcoholismo, violencia de género, entre otros. Más de setecientos profesionales registran y transcriben las consultas, clasificándolas internamente como demanda (D), problema (P) y recurso (R). Actualmente, se está implementando un algoritmo entrenado con trescientas mil entrevistas para sugerir los recursos adecuados para cada caso. Este esta ya en uso en tres centros. Adicionalmente, se está desarrollando un sistema de evaluación continua de servicios públicos

¹¹ https://ajuntament.barcelona.cat/premsa/wp-content/uploads/2021/04/Mesura-de-Govern-Intel·ligencia-artificial_cat-v2.47-ca-ES_.pdf

usando IA para comparar decisiones humanas con decisiones automatizadas, integrando un sistema de datos masivos para monitorear la evolución temporal y la efectividad de la ayuda social.

- *Monitorización de afluencia en las playas.* Área responsable: Barcelona Ciclo del Agua, SA. Durante la pandemia, se utilizó un sistema de procesamiento de imágenes y aprendizaje automático para monitorear la ocupación de las playas, comparando imágenes de las áreas llenas y vacías. El servidor activaba una alarma cuando era necesario y toda la información recogida se enviaba a una sala de decisiones, donde un equipo especializado en la monitorización de la capacidad de las playas decidía sobre posibles cierres. Esta información también se hacía accesible al público a través de un portal municipal, permitiendo a los ciudadanos consultar la capacidad de las playas en tiempo real.
- *Plan de robotización de la Estrategia de transición digital.* Área responsable: Acción transversal liderada por el Instituto Municipal de Informática, con la participación —entre otras áreas— de Instituto Municipal de Hacienda, el Distrito del Eixample y el Área de Ecología Urbana.
- *La Robotic Process Automation (RPA)* es una tecnología que, aunque no es IA, es fundamental en la transformación digital del Ayuntamiento de Barcelona. La RPA utiliza "robots" de software para emular e integrar acciones humanas en la ejecución de procesos, con el objetivo de aumentar la eficiencia robotizando tareas repetitivas anteriormente realizadas por personas. El plan de robotización ha establecido un nuevo servicio de RPA que facilita soporte a los robots y automatismos de diferentes áreas del Ayuntamiento, orquestando su actividad para proporcionar indicadores de uso y controlar la trazabilidad, asegurando su adecuación a requerimientos cambiantes. Ejemplos de uso incluyen la automatización de la tramitación de licencias de ocupación de espacio público para la diada de Sant Jordi en 2019 en el Distrito de l'Eixample, la gestión de expedientes para la ampliación de terrazas durante la pandemia en 2020 por Ecología Urbana, y la implantación de un piloto en el proceso de declaración de procedimientos concursales en 2020 por el Instituto Municipal de Hacienda, liberando recursos para tareas de mayor valor.
- *Mejoras en los canales de comunicación con la ciudadanía mediante IA.* Área responsable: Dirección de Servicios de Información y Atención Ciudadana. El Ayuntamiento de Barcelona, en colaboración con el Barcelona Supercomputing Center y el Instituto Municipal de Informática, está desarrollando un proyecto que utiliza inteligencia artificial para analizar los datos recogidos por el sistema IRIS, que gestiona incidencias y reclamaciones ciudadanas. El objetivo es identificar patrones en estas comunicaciones relacionados con eventos o situaciones específicas en la ciudad, lo cual

permitirá anticipar las necesidades de la ciudadanía y mejorar la respuesta municipal. Mediante algoritmos de procesamiento del lenguaje natural, se anonimizará la información antes de extraer datos de alto valor, clasificándolos por áreas temáticas y geográficas para encontrar correlaciones con eventos locales. Un resultado clave del proyecto será la creación de un conjunto de datos en catalán y castellano, especialmente en catalán donde actualmente hay escasez de datos. La principal fuente de datos será IRIS, con 260,000 comunicaciones anuales, y se considerará también la incorporación de datos del servicio telefónico de atención ciudadana, el 010, con unas ocho mil llamadas diarias.

ANEXO II. REFERENCIA A ALGUNOS ALGORITMOS PÚBLICOS

Aplicación en otras administraciones europeas:

- *Generalitat de Catalunya*. Tramitación de las becas universitarias. Equidad. Ayudas para cursar estudios de grado o máster en universidades públicas catalanas.
- Sistemas de aparcamientos públicos, servicios sanitarios municipales, mercado de alquiler local, servicios de maternidad y salud infantil. Sistema de recomendaciones de la biblioteca central y la gestión inteligente de la colección de libros y otros materiales (*Ayuntamiento de Helsinki*)
- Detección de fraudes en el alquiler vacacional, sistema de control de aparcamientos, tramite de quejas vecinales a partir del análisis del lenguaje natural (*Ayuntamiento de Amsterdam*)
- En el Reino Unido caben señalar los siguientes registrados¹²:
 - Asociación de Greater Cambridge:

Prueba de cruce de peatones inteligente: Un sensor de cámara con IA, llamado Starling, analiza en tiempo real el movimiento de peatones en cruces señalizados para mejorar la gestión del tráfico y la seguridad.

- Oficina del Gabinete:

Revisión automatizada de documentos digitales : Un algoritmo diseñado para identificar documentos digitales de valor histórico y aquellos sin valor a largo plazo.

Enlaces relacionados : Motor de recomendaciones que facilita la navegación en el portal GOV.UK, proporcionando rutas relevantes relacionadas con el contenido consultado.

- Agencia de Normas Alimentarias :

Esquema de calificación de higiene alimentaria – AI : Herramienta que asiste a las autoridades locales en la priorización de inspecciones sanitarias en establecimientos alimentarios, prediciendo los que presentan mayores riesgos.

- Policía de West Midlands :

¹² <https://www.gov.uk/algorithmic-transparency-records>

Análisis exploratorio de las condenas sexuales : Una herramienta que evalúa los factores que pueden influir en el éxito o fracaso de investigaciones relacionadas con delitos sexuales.

- Policía de Hampshire y Thames Valley :

DARAT (Herramienta de evaluación del riesgo de maltrato doméstico) : Asiste a los oficiales en la evaluación del riesgo de incidentes futuros de violencia doméstica.

- Oficina del Comisionado de Información:

Bandeja de entrada de registro AI : Automatiza la categorización y respuesta a correos electrónicos en la Oficina del Comisionado de Información, mejorando la eficiencia administrativa.

- Departamento de Salud y Asistencia Social y NHS Digital :

Algoritmo QCovid : Herramienta clínica para predecir el riesgo individual de complicaciones graves por coronavirus, apoyando las decisiones médicas.

Un caso interesante a seguir son las iniciativas establecidas en Chile, donde hay un registro exhaustivo de uso de IA en todas las administraciones y servicios para el ciudadano. Se han registrado los siguientes sistemas¹³:

¹³ <https://www.consejotransparencia.cl/wp-content/uploads/estudios/2021/10/ESTUDIO-TRANSPARENCIA-ALGORITMICA-EN-EL-SECTOR-PUBLICO-GOBLAB-cambio-tablas-1.pdf>

ANEXO III.FICHA EXTRAÍDA DE RADAR: “RADAR DE ALGORITMOS Y PROCESOS DE DECISIÓN AUTOMATIZADA PARA EL ACCESO A LOS DERECHOS SOCIALES DE LA CIUDADANÍA”

<i>Bloque 1. Información básica</i>		
Nombre	(En caso de tenerlo, nombre interno o público del algoritmo.)	
Descripción	Contexto en el cual se desarrolla y se aplica este proceso automatizado/ algoritmo/sistema de IA. Es decir, qué problemática se había detectado, cuál fue la solución por la que se optó, cómo ha sido la implementación y en qué consiste exactamente y qué resultados ha dado hasta ahora.	
Objetivo	¿Cuál es el objetivo del proceso automatizado/algoritmo/sistema de IA? ¿Para qué función ha sido desarrollado?	
Objetivo (Seleccionar objetivos que cumple el algoritmo. Opciones extraídas de Selected AI cases in the public sector, del Joint Research Centre de la Comisión Europea).	Automatización de tareas	Realizar de forma automatizada un conjunto de tareas que a una persona llevaría mucho más tiempo realizar
	Recopilación de datos personales	Recopilar de forma sistemática o predeterminada datos sobre individuos y/o grupos para fines públicamente conocidos o desconocidos y basándose en criterios públicamente conocidos o desconocidos.
	Evaluación del comportamiento humano	Generar evaluaciones de la forma como se comportan los individuos y/o grupos en base a criterios públicamente conocidos o desconocidos aplicados a datos públicamente conocidos o desconocidos

	<p>Reconocimiento de rasgos faciales</p>	<p>Identificar rasgos faciales particulares en imágenes de personas, como la forma de los ojos mientras una persona sonríe, basándose en criterios públicamente conocidos o desconocidos aplicados a datos públicamente conocidos o desconocidos.</p>
	<p>Identificación de imágenes de caras</p>	<p>Comparar imágenes de caras de personas individuales con imágenes de rostros registradas con anterioridad en una base de datos basándose en criterios públicamente conocidos o desconocidos aplicados a datos públicamente conocidos o desconocidos.</p>
	<p>Predicción del comportamiento humano</p>	<p>Generar posibles escenarios futuros donde individuos y/o grupos puedan comportarse en base a criterios públicamente conocidos o desconocidos aplicados a datos públicamente conocidos o desconocidos.</p>
	<p>Elaboración de perfiles y clasificación de personas</p>	<p>Generar perfiles de individuos y/o grupos y clasificarlos y ordenarlos en función de criterios públicamente conocidos o desconocidos aplicados a datos públicamente conocidos o desconocidos.</p>
	<p>Simulación del habla humana</p>	<p>Generar un habla que se parezca mucho a la forma como hablan las personas para propósitos públicamente conocidos o desconocidos</p>
	<p>Reconocimiento de imágenes</p>	<p>Identificar el contenido de imágenes digitales, por ejemplo, si es una imagen de un gato o de un perro, basándose en criterios públicamente conocidos o desconocidos aplicados a</p>

		datos públicamente conocidos o desconocidos.
	Generación de traducciones automáticas	Traducir automáticamente textos o discursos escritos de un idioma a otro/s.
	Generación de resultados de búsqueda en línea	Producir una lista ordenada de sitios web u otros recursos en línea en respuesta a una consulta de búsqueda, generalmente como términos de búsqueda escritos o hablados.
	Reconocimiento de sonidos	Identificar el contenido del habla u otros sonidos, por ejemplo, si se trata de una persona hablando o de un animal u objeto en particular, basándose en criterios públicamente conocidos o desconocidos aplicados a datos públicamente conocidos o desconocidos.
	Xatbot conversacional	Proporcionar de forma automatizada respuestas a las consultas ciudadanas.
<i>Bloque 2. Desarrollo e implementación</i>		
Administración responsable	Administración pública que está utilizando o utilizará el algoritmo.	
Entidad desarrolladora	Entidad que ha desarrollado el algoritmo a nivel técnico. Por entidad desarrolladora se hace referencia a la empresa o persona que ha creado el sistema de IA o proceso automatizado	
Estatus. ¿En qué situación se encuentra la implementación?	<ul style="list-style-type: none"> • En definición: se ha identificado una necesidad y se está definiendo la función (y diseño) del algoritmo. • Planificado: el sistema ya está diseñado, pero aún no se está desarrollando. • En desarrollo: el sistema se está desarrollando. • Piloto: el sistema está siendo probado. • Implementado: el sistema está implementado y en uso. • Ya no está en uso. 	

	<ul style="list-style-type: none"> • Otro: ...
Fecha de inicio	Si el algoritmo se está utilizando, año en que se empezó a utilizar. Si está en fase piloto, momento en que se inició la prueba. Si está planificado o en desarrollo, fecha estimada en que se empezará a utilizar
Fecha de finalización	En caso de que el algoritmo esté en desuso, ¿en qué año dejó de utilizar- se? Si está bajo prueba piloto, ¿cuándo finaliza la prueba?
<i>Bloque 3. Colectivo, beneficios y riesgos</i>	
Colectivo	Colectivo al cual está dirigido o sobre el cual tiene un mayor impacto directo el algoritmo. Por ejemplo, el algoritmo podría estar implementado en un proceso que únicamente concierne a menores de edad.
Beneficios	Principales beneficios, tanto para la Administración como para el ciudadano, obtenidos gracias a la implementación de este algoritmo/sistema de IA
Beneficios. Seleccionar los beneficios aportados. Las opciones están agrupadas en 3 categorías y están extraídas de Selected AI cases in the public sector, del Joint Research Centre de la Comisión Europea.	<ul style="list-style-type: none"> • Mejora de los servicios públicos <ul style="list-style-type: none"> ○ Servicios personalizados (la oferta de los servicios se ajusta mejor a las necesidades de la población) ○ Servicios públicos (centrados en el ciudadano: mayor accesibilidad del ciudadano a los servicios públicos) ○ Incrementar la calidad de la información y de los servicios públicos ○ Servicios públicos más sensibles, eficientes y rentables ○ Nuevos servicios o canales • Mejora de la eficiencia administrativa <ul style="list-style-type: none"> ○ Reducción de costes ○ Capacidad de respuesta de la operación gubernamental ○ Mejora de la gestión de los recursos públicos ○ Aumento de la calidad de procesos y sistemas ○ Mejor colaboración y mejor comunicación ○ Reducir o eliminar el riesgo de corrupción y abuso de la ley por parte de los funcionarios públicos ○ Permitir una mayor equidad, honestidad e igualdad • Capacidades de gobierno abierto <ul style="list-style-type: none"> ○ Mayor transparencia de las operaciones del sector público ○ Aumento de la participación pública en las acciones gubernamentales y en la formulación de políticas

	<ul style="list-style-type: none"> ○ Mejora del control público y la influencia en las acciones y políticas gubernamentales
<p>Riesgos. Seleccionar los riesgos que presenta o podría comportar el uso de este algoritmo/sistema de IA. Categorías extraídas del repositorio OASI, de Éticas.</p>	<ul style="list-style-type: none"> ● Discriminación de género ● Discriminación racial ● Discriminación religiosa ● Discriminación socioeconómica ● Otro tipo de discriminación ● Polarización social / radicalización ● Vigilancia estatal ● Amenaza a la privacidad ● Generar adicción ● Manipulación / cambios en el comportamiento ● Difundir información errónea ● Otro: ...
<p><i>Bloque 4. Supervisión y transparencia</i></p>	
<p>Intervención humana. ¿Qué rol juega la supervisión humana?</p>	<ul style="list-style-type: none"> ● El resultado del algoritmo es supervisado por una persona que tiene la <ul style="list-style-type: none"> ● decisión final ● El resultado del algoritmo es consultado por una persona, pero esta no <ul style="list-style-type: none"> ● puede modificar la decisión ● El resultado del algoritmo es supervisado/consultado por una persona, <ul style="list-style-type: none"> ● pero esta no tiene el conocimiento/capacidad para evaluarlo ● El resultado no se supervisa/consulta ● Otro: ...
<p>Transparencia. Grado de accesibilidad a la información de este algoritmo. ¿Existe algún documento público y/o página web donde se exponga la siguiente información?</p> <p>(Marque las que sí)</p>	<ul style="list-style-type: none"> ● Datos utilizados para entrenar el algoritmo ● Tipo de algoritmo utilizado ● Problemática a la cual se pretende dar solución ● Cómo se está implementando ● Objetivos y población que interactúa con el algoritmo ● Resultados obtenidos hasta ahora ● Otro: ...

<p>Visibilidad del ciudadano. ¿El ciudadano afectado directamente o indirectamente por la decisión/resultado del algoritmo es consciente de que se está usando este algoritmo?</p> <p>En otras palabras, no implica saber si el ciudadano efectivamente lo conoce, sino sobre si se ha hecho pública la información, esta es fácilmente accesible y está claramente visible</p>	<ul style="list-style-type: none">• Se informa directamente al ciudadano (comunicación activa)• La información es pública y accesible (comunicación pasiva)• La información es pública pero poco accesible (depende de la proactividad del ciudadano para informarse)• No se informa

ANEXO IV. LA GUÍA DE INFORMACIÓN ALGORÍTMICA PARA EL ÁMBITO LABORAL

Ahí se señala que se debe recoger la siguiente información de los algoritmos utilizados¹⁴:

A. Información general sobre el uso de algoritmos o sistemas de inteligencia artificial para tomar decisiones automatizadas

1. ¿La empresa utiliza o tiene previsto utilizar algoritmos y sistemas de decisión automatizada para la gestión de personas trabajadoras o candidatas? Sí
2. ¿La empresa utiliza algoritmos o sistemas de decisión automatizada para la elaboración de perfiles de las personas? Sí
3. ¿Qué decisiones de gestión de personas son automatizadas mediante algoritmos o sistemas de inteligencia artificial? Selección de personas
4. ¿Qué tipo de tecnología utiliza el algoritmo?
 - a) ¿Genera un modelo de “caja negra”? Sí
 - b) ¿Es un algoritmo de aprendizaje continuo? Sí
 - c) ¿Quién ha desarrollado el modelo?
5. ¿Cuál es el software o producto utilizado por el algoritmo?
 - a) ¿Quién es la empresa suministradora?
 - b) ¿Se han realizado modificaciones o alteraciones del software en su instalación? Sí
 - c) En su caso, ¿de qué tipo y con qué finalidad?
 - d) ¿El software cuenta con algún tipo de certificación? Sí
6. ¿Existe intervención humana cualificada en el proceso de decisión? Sí
 - a) En su caso, ¿cuál es el grado de intervención de la persona en la adopción de la decisión final?

14

https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/trabajo14/Documents/2022/100622-Guia_algoritmos.pdf

- b) ¿Cuál es la competencia y autorización de la persona que interviene en el proceso de decisión?
 - c) ¿Cuál es su capacidad para modificar la decisión adoptada por el algoritmo?
- B.** Información sobre la lógica y funcionamiento de cada algoritmo usado
7. En caso de elaboración de perfiles, ¿qué tipología de perfiles elabora el algoritmo?
- a) Específicamente para la persona trabajadora o candidata, ¿a qué perfil ha sido asignada?
 - b) El perfil ¿ha sido usado o está previsto usarlo para una finalidad distinta?
8. ¿Cuáles son las variables que utiliza el algoritmo?
- a) ¿Se han incluido como variables datos de carácter personal? Sí
 - b) ¿Qué tipología de datos de carácter personal son utilizados?
 - c) ¿Se han cumplido las obligaciones de información respecto del uso o, en su caso, reutilización de datos personales?
 - d) ¿Qué tipología de datos de carácter no personal son utilizados?
9. ¿Qué parámetros utiliza el algoritmo?
10. ¿Cuáles son las reglas e instrucciones que utiliza el algoritmo en el proceso de toma de decisión?
11. ¿Qué base de datos de entrenamiento y, en su caso, de validación se ha utilizado?
- a) ¿Los datos de entrenamiento son adecuados, pertinentes y no excesivos? Sí
 - b) ¿Se han verificado estos extremos? En caso afirmativo, ¿quién ha realizado la verificación?
 - c) ¿Su utilización como datos de entrenamiento se corresponde con la finalidad para la que fueron obtenidos? Sí
 - d) En su caso, ¿qué tipo de patrones se han identificado en los datos de entrenamiento?
12. ¿Se han detectado errores o imprecisiones del algoritmo en la clasificación de las personas en los diferentes perfiles? Sí
- a) ¿Cuál es el porcentaje aproximado de error?

- b) ¿Cuáles son las métricas de rendimiento por perfil?
13. ¿Se han realizado o se prevé realizar auditorías o evaluaciones de impacto? Sí
- c) ¿Se realizan con medios propios o externos? Propios
 - d) ¿Cuál ha sido el resultado de la auditoría y/o evaluación de impacto?
- C.** Información sobre las consecuencias que pueden derivarse de la decisión automatizada o del uso del algoritmo
14. ¿Qué consecuencias pueden derivarse para las personas trabajadoras o candidatas de la decisión adoptada por el algoritmo o del perfil creado?
15. A la representación legal de la plantilla, ¿cuál es el impacto el algoritmo en materia de igualdad y no discriminación entre mujeres y hombres?
- a) ¿Existe constancia de posibles problemas de sesgo? Sí
 - b) ¿Se ha incluido el impacto del algoritmo en el diagnóstico previo a la elaboración de un Plan de Igualdad? Si
- D.** Otra información relevante para la representación legal de la plantilla
16. ¿Se ha informado a las personas trabajadoras o candidatas respecto del uso de algoritmos para la toma de decisiones automatizadas? Sí

ANEXO V. FICHA DE TRANSPARENCIA ALGORÍTMICA DE LA AOC EN CATALUÑA

AOC (Administración abierta de Cataluña)¹⁵ en su “Fitxa de transparència algorítmica”¹⁶

La AOC propone publicar fichas comprensibles para cada algoritmo de IA en el Portal de Transparencia. Estas fichas deben incluir:

- Problema que resuelve el algoritmo.
- Características y funcionamiento del algoritmo.
- Colectivos afectados.
- Evaluación de impacto y posibles sesgos.
- Análisis de riesgos de seguridad.
- Medidas de gestión de riesgos para garantizar derechos fundamentales.
- Órganos de recurso contra decisiones algorítmicas.
- Datos de contacto del responsable y proveedor.

Las fichas deben identificar y analizar los riesgos relacionados con derechos fundamentales, proponiendo medidas adecuadas para su mitigación, asegurando así el cumplimiento de los derechos.

TÍTULO		
VISIÓN GENERAL	Casos de uso	
	Estatus del servicio	En definición Piloto Implementado Otro
	Acuerdos de nivel de servicio	
	Beneficios del servicio	
	Colaboración	
Información de contacto		
Información más detallada del servicio	Conjunto de datos	Conjunto de datos 1 Conjunto de datos 2 Consideraciones sobre el tratamiento y conservación de los datos

¹⁵ <https://www.aoc.cat/es/projecte-innovacio/transparencia-en-lus-dalgorismes-dintelligencia-artificial-a-laoc/>

¹⁶ https://www.aoc.cat/wp-content/uploads/2023/05/FIT_TRANSP_ALGORISMICA_ESTRUCTURA_web.pdf

	Procesamiento de los datos	Proceso paso a paso (incluir el tipo de algoritmo utilizado) Arquitectura del sistema Rendimiento del algoritmo
	Supervisión humana	
	Cumplimiento normativo del sistema	¿Es una actuación administrativa automatizada AAA?
	Gestión de los riesgos (y medidas aplicadas para garantizar el cumplimiento de los derechos fundamentales)	<ol style="list-style-type: none"> 1. Igualdad y no discriminación. 2. Protección de datos y privatización 3. Seguridad y robustez 4. Transparencia y explicabilidad 5. Rendimiento de cuentas y auditabilidad 6. Desarrollo sostenible
Más información	Bibliografía y enlaces de interés	

ANEXO VI. TABLA RESUMEN COMPARADO DE LA INFORMACIÓN PROPORCIONADA EN LOS REGISTROS O ESTÁNDARES ESTABLECIDOS O PROPUESTOS

ALGORITHMIC TRANSPARENCY STANDARD	ICO	ÁMSTERDAM Y HELSINKI	PAÍSES BAJOS	OBSERVATORIO IA EN SALUD (CATALUÑA)	GUÍA ETALAB FRANCIA
INFORMACIÓN BÁSICA	INFORMACIÓN GENERAL	VISIÓN GENERAL	INFORMACIÓN GENERAL	DATOS BÁSICOS DEL ALGORITMO	SOBRE LA ADMINISTRACIÓN
Nombre del algoritmo	Nombre del algoritmo	Objetivo principal al que contribuye el algoritmo	Nombre el algoritmo	Nombre del algoritmo	Nombre de la administración responsable
Descripción corta	Descripción breve	¿Cómo se está utilizando?	Organización responsable	Procedimiento en el que se enfoca el algoritmo	Dirección/servicio
Organización responsable	Página web	¿En qué casos se utiliza?	Departamento	Qué tecnologías utiliza el algoritmo	Contacto
Departamento	E-mail de contacto	¿Sobre quién tiene impacto?	Descripción corta	Responsables del algoritmo	Fecha de actualización de la información
E-mail de contacto	PROPIEDAD Y RESPONSABILIDAD	¿Cuáles son los impactos esperados?	Tipo de algoritmo	Correo electrónico de los responsables	ALGORITMO Y DECISIÓN TOMADA
Número de teléfono de contacto	Administración	¿Cómo funciona el sistema a un alto nivel?	Dominio en el que se despliega el algoritmo	Teléfono de contacto de los responsables	Nombre del algoritmo
Nombre de contacto	Órgano y, en su caso, servicio o equipo concreto responsable	RENDICIÓN DE CUENTAS	Enlace a web de publicación	¿Algoritmo asociado a algún proyecto?	Contexto global
Área geográfica	Persona responsable dentro del órgano	Organizaciones responsables	Estado	Desarrolladores del algoritmo	Finalidad algoritmo
Dominio	Participación de proveedores externos e identificación	Departamentos responsables	APLICACIÓN	Correo electrónico de los desarrolladores	Cómo interviene el algoritmo en la decisión
Página web	Nivel de intervención del proveedor externo	Persona, email y teléfono de contacto	Objetivo	Teléfono de contacto de los desarrolladores	Cuando interviene el algoritmo en la decisión

Estado	Tipo de procedimiento de contratación pública	Socios y proveedores externos que intervienen	Impacto	¿Otra entidad involucrada?	Nivel de automatización de la decisión
Observaciones	Régimen de acceso a datos de los proveedores externos	CONJUNTO DE DATOS	Proporcionalidad	¿Algún centro de investigación involucrado?	Fundamentos jurídicos
ÁMBITO DE APLICACIÓN Y CONTROL	DESCRIPCIÓN Y JUSTIFICACIÓN	Nombre del conjunto de datos	Proceso de toma de decisiones	Propiedad del algoritmo	Recursos relacionados
Proceso de toma de decisiones	Descripción detallada	Descripción de la información representada en el conjunto de datos	LEGAL	Propiedad intelectual del algoritmo	IMPACTO DE LA DECISIÓN
Objetivos	Ámbito de aplicación	PROCESAMIENTO DE DATOS	Organismo autorizado para el despliegue del algoritmo	Derechos por el uso del algoritmo	Número de decisiones administrativas tomadas por año
Impacto	Beneficios de la utilización del sistema	Arquitectura modelo del sistema	Base legal	¿Algún proveedor externo involucrado?	Alcance de la decisión
Riesgos y medidas	Explicación de cómo se llevaba el proceso a cabo anteriormente	Comportamiento de sistema	Evaluación de impacto en los derechos humanos (IAMA)	ÁMBITO DE APLICACIÓN, OBJETIVOS E IMPACTO	Público afectado por la decisión
Categoría de riesgo (nivel de riesgo)	Alternativas consideradas	NO DISCRIMINACIÓN	Evaluación de impacto de protección de datos (DPIA)	Ámbito de actuación del algoritmo	Recursos relacionados
Proporcionalidad	PROCEDIMIENTO DE TOMA DE DECISIONES	¿Cómo se interpreta el sesgo injusto en el contexto del sistema?	Procedimiento de reclamación	Objetivo principal	FUNCIONAMIENTO INTERNO DEL ALGORITMO
Base legal	Forma en la que el algoritmo se integra en el procedimiento	¿Qué medidas existen para comprobar el sesgo?	METADATOS	Objetivos específicos	Datos tratados
Más información	Información proporcionada por el algoritmo en el marco del procedimiento	¿Es el sistema diseñado e implantado teniendo en cuenta la accesibilidad del servicio para personas	Enlace a la web del registro de origen	Nivel de automatización	Fuente de los datos

		con discapacidades?			
DATOS	Nivel de intervención y supervisión humana	Implicación de las personas afectadas por el sistema en el diseño y desarrollo	E-mail de contacto	ÁMBITO DE APLICACIÓN	Modo de recolección de los datos
Datos de entrenamiento	Formación específica proporcionada las personas que emplean el sistema	SUPERVISIÓN HUMANA	Área geográfica	Descripción del uso del algoritmo	Tipo de algoritmo
Fuente de los datos	Sistema de recursos, revisión y quejas	Capacidad y soporte para la intervención humana en el diseño y desarrollo del sistema, ciclos de decisión, y en el seguimiento del funcionamiento del sistema	Idioma	Explicación a alto nivel del funcionamiento tecnológico del sistema	Operaciones efectuadas por el algoritmo
Sesgo de los datos	ESPECIFICACIONES TÉCNICAS Y DATOS	Competencias necesarias para la supervisión	ÁMBITO DE APLICACIÓN	Especialidad médica a la que se aplica el algoritmo	Recursos relacionados
Conexión con otras fuentes	Método usado por el sistema	Entrenamiento disponible para adquirir las competencias necesarias	Descripción larga	¿Algoritmo asociado a algún diagnóstico?	
Descripción de la evaluación de impacto de la protección de datos (DPIA)	Frecuencia de uso	RIESGOS	Uso de los registros base	Colectivo beneficiado	
Enlace al DPIA	Fase de desarrollo	Compensaciones clave entre riesgos y beneficios	Fuentes de datos	NIVEL DE MADUREZ DEL ALGORITMO	

ALGORITMO	Actuaciones necesarias para su mantenimiento	Nivel de riesgo	Métodos y modelos utilizados por el algoritmo	Fase del algoritmo	
Tipo de algoritmo	Rendimiento del sistema: incluyendo precisión, número de veces que una persona se ha apartado de la decisión recomendada y recursos contra las decisiones del sistema que han sido estimados	Métodos de gestión de los riesgos	SUPERVISIÓN	Nivel de madurez del algoritmo	
Explicación del funcionamiento	Arquitectura del sistema	Riesgos identificados	Monitorización	¿Está implementado el algoritmo?	
Métodos y modelos	Fuente de los datos de entrenamiento y funcionamiento	Medidas implementadas para mitigar los riesgos	Intervención humana	¿Es un proyecto en activo?	
Comportamiento del algoritmo	Descripción de los datos de entrenamiento y funcionamiento	Evaluaciones de impacto en la privacidad	Riesgos	¿Cuándo se ha iniciado?	
Alternativas consideradas	En su caso, página web donde se encuentran los datos	Evaluaciones de impacto en derechos humanos			
Enlace a la información central del producto (el algoritmo)	Forma en que se recogieron los datos	Otra evaluación de riesgos asociados a los métodos aplicados			
Enlace al código fuente	Procesos de limpieza de los datos	EXPLICABILIDAD			
Enlace a la página del producto	Análisis del nivel de integridad y representatividad de los datos	Razones que llevaron al resultado concreto con la aplicación del algoritmo			
Evaluación del impacto del algoritmo	Acceso y almacenamiento de datos por parte del proveedor del sistema	REFERENCIAS			

Enlace a la evaluación del impacto del algoritmo	RIESGOS MITIGACIÓN I EVALUACIONES DE IMPACTO	Referencias a documentación adicional : políticas de privacidad, contratos, informes de auditoría...			
SUPERVISIÓN	Evaluación de impacto: resumen y enlace de acceso a la evaluación completa				
Monitorización del comportamiento	Explicación de los riesgos y estrategias de mitigación				
Tipo de uso del algoritmo					
Intervención humana					
Procedimientos de reclamación					
Retroceso (posibilidad de anular los efectos de la aplicación del algoritmo)					
Metadatos					
Versión del estándar utilizado					
Identificador único					
Idioma					
Fecha de publicación					
Fecha de la última actualización					
Fecha de revisión del registro					
Nota sobre la revisión (revisión ordinaria, cambios en la ley...)					
Palabras clave					

