



Taula d'entitats  
del Tercer Sector Social  
de Catalunya

m4social

INFORME

**RADAR DE ALGORITMOS  
Y PROCESOS DE DECISIÓN  
AUTOMATIZADA PARA EL ACCESO  
A LOS DERECHOS SOCIALES  
DE LA CIUDADANÍA**

ENERO DE 2024

# CRÉDITOS

**Estudio impulsado por:**

Taula d'entitats del Tercer Sector Social de Catalunya

**Coordinación:**

Roger Civit  
Laura Castro  
Xènia Carné  
Gerard Sentís

**Autoría del estudio [KSNET]:**

Romén Adán González  
Mireia Álvarez  
Elena Costas

**Peer reviewers [KSNET]:**

Beatriz Belmonte Acero  
Manuel González Noriega

**Diseño:** Toni Sánchez Poy

**Licencia:** CC BY-NC-SA 4.0



Barcelona, febrero 2024

# ÍNDEX

Resumen.....	4
--------------	---

## Documento 1. La inteligencia artificial en las administraciones públicas

1. <b>Introducción</b> .....	7
2. <b>Aclaración de conceptos</b> .....	10
2.1 Conceptos clave.....	10
2.2 Un ejemplo de la aplicación.....	12
3. <b>El origen de la inteligencia artificial y los últimos avances</b> .....	16
4. <b>Gobernanza y estrategias nacionales en el ámbito de la inteligencia artificial</b> .....	21
4.1 El rol de los estados en la gobernanza de la IA.....	23
4.2 El rol de los estados en la aplicación de la IA en el sector público.....	24
4.3 La estrategia nacional de España.....	25
5. <b>El enfoque centrado en la persona y la ley de inteligencia artificial de la Unión Europea</b> .....	27
5.1 El enfoque centrado en la persona.....	27
5.2 La ley de IA de la Unión Europea.....	30
5.3 Las acciones de España.....	34
<i>El Real decreto de sandbox de España</i> .....	34
<i>Estatuto de la agencia española de supervisión de la Inteligencia Artificial (AESIA)</i> .....	35
6. <b>La inteligencia artificial en las administraciones públicas</b> .....	36
6.1 ¿Qué entendemos por IA en el contexto de las AAPP?.....	36
6.2 Aplicaciones de la IA en las AAPP.....	37
6.3 Casos internacionales.....	43
6.4 Aplicaciones en el ámbito social.....	48
7. <b>Beneficios y riesgos de la aplicación de la inteligencia artificial en las administraciones públicas en el ámbito social</b> .....	51
8. <b>Conclusión</b> .....	56
<b>Referencias</b> .....	57

## Documento 2. Radar de algoritmos y procesos de decisión automatizada para el acceso a los derechos sociales de la ciudadanía

1. <b>Introducción</b> .....	63
2. <b>Metodología</b> .....	65
3. <b>El diseño del Radar de algoritmos</b> .....	67
3.1. Campos del Radar de algoritmos.....	67
4. <b>Resultados</b> .....	74
Fichas informativas.....	78
1. <i>Inteligencia colectiva - sistema de identificación de demandas, problemas y respuestas (DPR automático)</i> .....	78
2. <i>Procesador de casos IRIS</i> .....	82
3. <i>Sistema de apoyo a la clasificación de incidencias - MARIO</i> .....	83
4. <i>Sistemas conversacionales, infoCanal Alzheimer y fondo 0-16</i> .....	84
5. <i>Transcripción de informes de emergencia social (CUESB)</i> .....	86
6. <i>RisCanvi</i> .....	87
7. <i>Motor de Reglas de Prestaciones Sociales</i> .....	90
8. <i>Recomendador de prestaciones</i> .....	91
9. <i>Identificación necesidades en la Intervención Social</i> .....	92
10. <i>Xatbots conversacionales de la Administración Abierta de Cataluña (AOC)</i> .....	94
11. <i>Servicio de videoidentificación para obtener el idCAT móvil</i> .....	98
12. <i>Servicio de automatización de los informes de pobreza energética</i> .....	102
5. <b>Valoraciones finales</b> .....	108

# RESUMEN

En la última década, la inteligencia artificial (IA) se ha desarrollado de forma exponencial. Una evolución que ha sido principalmente impulsada por los progresos en el campo del aprendizaje profundo (*deep learning*). La IA generativa, tanto de texto como de imagen o audio, se encuentra hoy en un nivel de desarrollo difícilmente imaginable pocos años atrás. Por otro lado, los sistemas que utilizan IA están cada vez más integrados en la vida cotidiana. Desde asistentes personales inteligentes hasta sistemas de recomendación en plataformas, *bots*, modelos que ayudan en el diagnóstico y tratamiento de enfermedades, o en la traducción y transcripción de textos.

Además, **las posibles aplicaciones de la IA poseen un potencial transformador, no solo en el sector privado, sino también en el sector público y, en particular, en el ámbito social.** La capacidad para simplificar procesos o proveer mejores servicios públicos puede facilitar y agilizar tanto el trabajo del personal de la Administración como la relación de los ciudadanos con esta. No obstante, el uso y aplicación de sistemas de IA, además de beneficios, comporta igualmente una serie de riesgos.

En este contexto, **desde la Taula d'entitats del Tercer Sector Social se ha impulsado el proyecto *Radar de algoritmos de inteligencia artificial y procesos de decisión automatizada para el acceso a derechos sociales de la ciudadanía de Cataluña*.** Este proyecto cuenta con un doble objetivo: por un lado, ayudar a la sociedad en general, y a las entidades del tercer sector en particular, a entender el contexto actual para poder fijar una posición compartida. En segundo lugar, mapear todos aquellos sistemas de inteligencia artificial o de decisión automatizada que están siendo utilizados actualmente por parte de las administraciones públicas catalanas en el ámbito social.

Por esta razón, **el presente trabajo se compone de dos documentos.** El primero de ellos, ***La inteligencia artificial en las administraciones públicas***, entregado en septiembre de 2023, es un informe de carácter divulgativo sobre los orígenes, evolución, gobernanza, beneficios, riesgos y aplicación de la inteligencia artificial por parte de la Administración pública.

El segundo documento, ***Radar de algoritmos y procesos de decisión automatizada para el acceso a los derechos sociales de la ciudadanía***, finalizado en diciembre de 2023, reúne un total de 12 sistemas identificados de 4 administraciones públicas catalanas diferentes. Para cada uno de estos algoritmos se ha elaborado una ficha informativa que recoge información relevante y variada. Este radar es el primer repositorio de sistemas de inteligencia artificial y procesos automatizados en el ámbito social utilizados por administraciones públicas de Cataluña que se ha creado a día de hoy.

En un contexto de profundas transformaciones es fundamental poder identificar lo que es relevante, los avances y riesgos que comporta, y cómo y qué se espera conseguir de lo que se está desarrollando. Con este trabajo se busca contribuir a la divulgación sobre inteligencia artificial, así como a la discusión pública, principalmente dando una mayor difusión pública a aquellos sistemas de inteligencia artificial que están siendo utilizados por parte de la Administración.

DOCUMENTO 1

# LA INTELIGENCIA ARTIFICIAL EN LAS ADMINISTRACIONES PÚBLICAS

SEPTIEMBRE DE 2023

# 1. INTRODUCCIÓN

Inteligencia artificial (IA) se ha convertido en un término tan popular como (quizás) poco útil. Un concepto paraguas que en ocasiones cuesta averiguar a qué está haciendo referencia cuando se utiliza. Si bien hace años que convivimos con los avances científicos en el campo de la IA, con su uso en el sector privado y su integración en muchas de las aplicaciones y webs que utilizamos en el día a día, hasta la actualidad los medios de comunicación y la opinión pública aún no habían puesto especial atención en ella. Probablemente todo ha cambiado con la aparición de ChatGPT, una interfaz amigable que permite a cualquier persona con acceso a internet interactuar directamente y conscientemente con un sistema de inteligencia artificial. Un sistema que, en pocos segundos, produce respuestas similares a las que podría proporcionar una persona.

Eso ha generado muchas dudas en el debate público. ¿La mayoría de personas perderán su trabajo? ¿Acabará la IA con el pensamiento crítico? ¿O incluso con la humanidad? Sin embargo, es importante generar espacios de debate sosegados sobre la inteligencia artificial. **De qué hablamos exactamente cuando hablamos de IA? ¿Cuáles son sus orígenes y cuál ha sido su desarrollo? ¿Qué uso se le ha dado en el sector privado en la última década? ¿Cómo podría beneficiarse la Administración pública y qué aplicaciones innovadoras se están desarrollando en otros países? ¿Qué beneficios puede aportar la IA a la sociedad? ¿Qué riesgos existen y cómo debemos afrontarlos y regularlos?, etc.**

El proyecto *AI Impacts* realizó en 2022 la segunda edición de la *Expert Survey on Progress in AI*<sup>1</sup>, donde se entrevistaron a 738 investigadores expertos en *machine learning* sobre los impactos de la inteligencia artificial en el futuro. Aunque no deja de ser un juicio valorativo, la mitad de los expertos en inteligencia artificial cree que la probabilidad de que el efecto a largo plazo de la IA avanzada en la humanidad sea «extremadamente malo (por ejemplo, la extinción humana)» es solo del 5 % o menos. Es más, uno de cada cuatro expertos la reduce a un 0 %.

Si bien la inteligencia artificial tiene un potencial de cambio muy destacado en el sector privado, el sector público no está exento de este. Diversas administraciones públicas ya cuentan con esta tecnología para ayudarles en diferentes tareas. De hecho, cabe señalar que, actualmente, la principal tecnología utilizada por el sector público son los sistemas de decisión automatizada (o ADMS, por sus siglas en inglés), que pueden (o no) utilizar IA.

<sup>1</sup> Puede consultarse [aquí](#).

Los sistemas de decisión automatizada (ADMS) son aquellos que tienen capacidad para tomar decisiones a través de medios tecnológicos sin la intervención humana. Pueden estar basados en cualquier tipo de datos, como aquellos facilitados por los individuos a través de cuestionarios, datos observados como los de geolocalización o datos inferidos a partir de unas ciertas características<sup>2</sup>. Como se explica en el informe *Automating Society* d'AlgorithmWatch<sup>3</sup>, los ADMS pueden (o no) utilizar técnicas de inteligencia artificial, es decir, pueden recurrir a procedimientos simples de análisis basados en reglas y automatizaciones, o implementar técnicas sofisticadas como el procesamiento de lenguaje natural, la analítica predictiva o la visión por computadora. En cualquier caso, los autores abogan por adoptar un enfoque holístico, de ahí el uso de la palabra sistemas en lugar de tecnologías. Al hacer referencia a un sistema, no se toma en consideración únicamente el algoritmo o técnica de forma aislada, sino todo el proceso de toma de decisión, el diagnóstico que explica por qué se adoptó esa solución, el algoritmo en sí, los datos, el código y la forma en la que fue desarrollado (por parte de una compañía pública o privada) y finalmente implementado. Desde un enfoque holístico puede comprenderse el proceso integral, desde el momento en que se identifica una problemática o necesidad hasta que finalmente se implementa una solución, pasando por todos los pasos intermedios necesarios.

Así pues, en este informe se utiliza el concepto de IA como un concepto amplio que incluye tanto los ADMS como la inteligencia artificial, dado que actualmente los ADMS se encuentran con mayor frecuencia dentro del sector público, y presentan oportunidades y riesgos similares a los de la IA.

Al situar la IA como un recurso del sector público a la hora de, por ejemplo, asignar ayudas sociales entre la ciudadanía, es imprescindible que las entidades del tercer sector conozcan cómo se están utilizando estos sistemas en el ámbito de los derechos sociales. Así pues, **la finalidad de este informe es ayudar a la sociedad en general, y a las entidades del tercer sector en particular, a entender el contexto actual para poder fijar una posición compartida**. Si las entidades quieren ser escuchadas, es fundamental que compartan y transmitan una visión similar, tanto en términos de diagnóstico como de propuestas.

Bajo este punto de partida, el informe se estructura de la siguiente forma: en el segundo apartado se definen brevemente una serie de conceptos

---

<sup>2</sup> Esta definición se ha extraído del documento *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679*, elaborado por el Data Protection Working Party, un órgano asesor de la Unión Europea. Puede consultarse [aquí](#).

<sup>3</sup> Disponible [aquí](#).

que aparecen a lo largo del informe. En el tercer capítulo se expone el origen y desarrollo de la IA hasta la actualidad. A continuación, se presenta un breve ejemplo para aclarar el concepto de entrenamiento de una IA. En los apartados 4 y 5 se pone el foco en la gobernanza de la IA, el papel de los gobiernos y la Ley de IA de la Unión Europea. El capítulo 6 presenta aplicaciones de sistemas que utilizan IA por parte de administraciones públicas (AA. PP.), tanto internacionales como en Cataluña y España, en diversos ámbitos. Finalmente, el capítulo 7 hace un repaso sobre los principales beneficios y, sobre todo, riesgos de la aplicación de la IA por parte de las AA. PP. en el ámbito social.

## 2. ACLARACIÓN DE CONCEPTOS

La inteligencia artificial (IA) es un campo amplio y complejo, con muchos subapartados y técnicas asociadas. Este concepto paraguas aglutina una amplia diversidad de técnicas: aprendizaje automático (*machine learning*), procesamiento de lenguaje natural, reconocimiento de voz, sistemas expertos, robótica o visión por computadora. Antes de empezar, para evitar confusiones, a continuación se describen algunos de los conceptos utilizados a lo largo del informe.

### 2.1 Conceptos clave

- ▮ **Algoritmo:** en términos generales, un algoritmo es un conjunto de reglas o instrucciones que, paso a paso, resuelven un problema o consiguen un objetivo. En el contexto de la IA, los algoritmos a menudo se refieren a los métodos utilizados para construir y entrenar modelos de aprendizaje automático. Existe una gran cantidad de algoritmos diferentes, y, a la hora de crear un modelo de IA, el científico de datos escoge el más adecuado para el problema a resolver.
- ▮ **Datos:** sin datos no hay inteligencia artificial. Los datos son la fuente de información. Buena parte de la creación de un modelo se centra en la recolección y limpieza de los datos. Los datos pueden ser imágenes, texto o una matriz de datos (como una hoja de Excel). Independientemente de su formato original, la información siempre se convierte en números para que el algoritmo escogido pueda aprender.
- ▮ **Entrenamiento:** es la fase de creación del modelo. Después de haber recogido y procesado una cantidad de datos determinada, estos son pasados a un algoritmo concreto para que «aprenda» a reconocer los patrones. Por ejemplo, si se dispone de información sobre la ubicación, las características y el precio de viviendas, se podría entrenar un modelo para que aprenda a predecir el precio de una vivienda nueva a partir de su ubicación y características.
- ▮ **Inteligencia artificial (IA):** no existe consenso en la comunidad científica sobre cómo definir el término IA. No obstante, la Comisión Europea ha propuesto una definición operacional que permite discernir responsabilidades legales, y que está basada en una concepción de la OCDE. El artículo 3(1) de la propuesta de Ley de Inteligencia Artificial establece que un «sistema de inteligencia artificial» implica:

«Un software que es desarrollado con técnicas y métodos específicos y que puede, para un conjunto dado de objetivos definidos por un humano, generar salidas como contenido, predicciones, recomendaciones o decisiones que influyen en el entorno con el que interactúa.»

El listado de técnicas figura en el anexo 1 de la propuesta de Ley de Inteligencia Artificial e incluye:

a) *Enfoques de aprendizaje automático, incluido el aprendizaje supervisado, no supervisado y de refuerzo, utilizando una amplia variedad de métodos, incluido el aprendizaje profundo;*

b) *Enfoques basados en la lógica y el conocimiento, incluida la representación del conocimiento, la programación (lógica) inductiva, las bases de conocimiento, los motores de inferencia y deducción, el razonamiento (simbólico) y los sistemas expertos;*

c) *Enfoques estadísticos, estimación bayesiana, métodos de búsqueda y optimización.*

► **Machine learning (aprendizaje automático):** es un subcampo de la IA que se centra en el diseño de sistemas que pueden aprender de los datos. En otras palabras, en lugar de programar directamente todas las acciones que realizará el modelo (como sucedía en los años 60 y 70), los sistemas de aprendizaje automático se entrenan con un conjunto de datos, aprendiendo patrones que después se utilizan para hacer predicciones o clasificaciones. Se consideran algoritmos de *machine learning* los algoritmos de regresión lineal, regresión logística, *k-nearest neighbours*, árboles de clasificación o el *random forest*, entre muchos otros.

► **Deep learning (aprendizaje profundo):** es un subcampo del aprendizaje automático que se centra en las redes neuronales artificiales con muchas capas (o «profundas»). Las redes neuronales son otro tipo de algoritmo dentro del campo de la IA. Estos modelos son capaces de aprender patrones complejos y sutiles en los datos. Se utilizan en una amplia variedad de aplicaciones, incluyendo el reconocimiento de voz, la visión por computadora y el procesamiento del lenguaje natural.

► **Algoritmos de clasificación y predicción:** los términos *clasificación* y *predicción* se utilizan en el campo del aprendizaje automático para describir dos tipos diferentes de problemas que estos algoritmos pueden abordar.

► Los **algoritmos de clasificación** se utilizan para predecir/identificar la categoría o la clase de un objeto o un evento. Un ejemplo de clasificación podría ser la detección de correo basura, donde se clasifican los correos electrónicos en «correo basura» o «no correo basura». O, por

ejemplo, un modelo que decide si a una persona le corresponde o no una ayuda social. Existen diferentes tipos de algoritmos de clasificación, como la regresión logística, los árboles de decisión o el vecino más próximo (a menudo abreviado k-NN).

- Los **algoritmos de predicción** se centran en predecir un valor continuo. Por ejemplo, pueden utilizarse para predecir el precio de una vivienda dadas sus características, o cuánto subirá la inflación el próximo mes. En este conjunto encontramos algoritmos como la regresión lineal o los árboles de regresión.

▫ **IA generativa o modelos fundacionales:** hace referencia a los sistemas de IA que pueden crear contenido nuevo y original. Eso puede abarcar desde la generación de texto y la creación de imágenes hasta la música y otros tipos de contenido. El modelo GPT-4 de OpenAI, que está detrás de ChatGPT, es un ejemplo de IA generativa en el campo del texto.

▫ **IA simbólica:** hace referencia a los desarrollos de la primera ola (años 60 y 70). La IA simbólica se refiere a los enfoques para desarrollar máquinas inteligentes mediante la codificación de los conocimientos y la experiencia de los expertos en conjuntos de reglas que pueden ser ejecutadas por la máquina.

## 2.2 Un ejemplo de la aplicación

Una vez definidos los conceptos que se utilizarán a lo largo de este informe, y para acabar de entender cómo funciona esta tecnología, a continuación ponemos un ejemplo sencillo e hipotético que nos sirva para tener una intuición sobre la mayoría de sistemas que utilizan algoritmos de *machine learning*, respondiendo a la siguiente pregunta: ¿Cómo se entrenan estos modelos de aprendizaje automático o *machine learning*?

Supongamos que una entidad del tercer sector lleva a cabo un programa de clases de refuerzo de matemáticas y lengua para alumnos en situaciones socioeconómicas complejas. Colaboran con más de 100 institutos, a los cuales envían, en un primer momento, un único profesor de refuerzo. Cuando se considera necesario, la entidad envía un nuevo profesor a un instituto para poder mejorar la calidad y el impacto del proyecto.

No obstante, el proceso para decidir si enviar o no un profesor extra de refuerzo es costoso y comporta un tiempo considerable. El profesor del proyecto debe rellenar un informe, expertos de la entidad deben acudir al instituto y evaluar el caso, etc. Con la experiencia, la entidad ha detectado que el proceso es tan costoso que no están siendo capaces de dar res-

puestas a tiempo. Llegados a este punto, deciden apoyarse en un sistema de inteligencia artificial.

Afortunadamente, el equipo de datos de la entidad ha ido recopilando información de los diferentes proyectos, anotando el municipio donde se encuentra el instituto, la renta per cápita del barrio, la nota media de los alumnos o la tasa de abandono escolar. Además, han incluido una columna en su hoja de Excel que indica si se ha enviado o no un profesor de refuerzo extra al proyecto.

Proyecto	Municipio del instituto	Renta per cápita del barrio del instituto	Nota media de los alumnos	Tasa de abandono escolar del instituto	Profesor de refuerzo extra
1	A	25	7	7	No
2	B	10	5,6	43	Sí
3	B	8	8	33	Sí
4	C	17	7,2	20	No
...	...	...	...	...	...
50	D	16	7,5	15	No

A partir de esta información, el equipo de datos decide entrenar un modelo de *machine learning* que prediga si es necesario o no asignar un profesor extra de refuerzo a un proyecto. Los analistas de datos no decidirán qué criterio seguir, sino que dejarán que sea el modelo quien «aprenda» de los datos. Es decir, el modelo identificará bajo qué circunstancias pasadas la entidad decidió enviar un nuevo profesor de refuerzo. Este proceso de aprendizaje es lo que se denomina entrenamiento del modelo, de forma que el modelo aprenderá de los siguientes datos:

Variables explicativas				Outcome
Municipio del instituto	Renta per cápita del barrio del instituto	Nota media de los alumnos	Tasa de abandono escolar del instituto	Profesor de refuerzo extra
A	25	7	7	No
B	10	5,6	43	Sí
B	8	8	33	Sí
C	17	7,2	20	No
...	...	...	...	...
D	16	7,5	15	No

La idea del entrenamiento radica en que el algoritmo aprenda a predecir el resultado (*outcome*) más probable a partir de las variables explicativas con las que cuenta. Por ejemplo, los proyectos de unos determinados municipios, con una tasa de abandono por debajo de cierto valor y una nota media concreta, tienden a tener asignado un profesor de refuerzo.

¿Cómo saber si el algoritmo da un buen resultado? El equipo de datos comparará los resultados reales (si el proyecto tiene o no asignado un profesor de refuerzo) con los estimados por el algoritmo para cada proyecto. Si el modelo apenas acierta un 30 % de los casos, el modelo será descartado, ya que no aporta ninguna utilidad a la entidad. El equipo de datos probará con diferentes modelos y ajustes hasta dar con aquel que consiga un mayor grado de acierto.

Una vez entrenado el modelo, el equipo de datos lo aplica sobre los restantes 50 proyectos para los cuales no saben si enviar o no un profesor de refuerzo. El modelo ha estimado que 5 institutos necesitarían un profesor de refuerzo. Ahora bien, la entidad, en lugar de haber analizado los 50 proyectos caso por caso para valorar en cuál sería necesario contar con un profesor extra, se centrará en los 5 identificados por el modelo como prioritarios.

Proyecto	Variables explicativas				Outcome
	Municipio del instituto	Renta per cápita del barrio del instituto	Nota media de los alumnos	Tasa de abandono escolar del instituto	Profesor de refuerzo extra (estimado)
51	B	22	7,4	12	No
52	A	8	5	31	Sí
...	D	...	...	...	...
99	D	12	5,2	6	Sí
100	C	25	6,5	25	No

Es importante destacar que según qué algoritmo de *machine learning* se haya utilizado, el equipo de datos sabrá explicar por qué el modelo estima que se necesita un profesor de refuerzo o no. Los modelos más precisos, es decir, aquellos que más aciertan, suelen ser más opacos, «cajas negras», de forma que el equipo de datos tendrá dificultades para explicar a la entidad el porqué del resultado. Por otro lado, los algoritmos más sencillos son más explicables, pero pueden cometer muchos más errores.

En conclusión, ¿cuál ha sido el proceso seguido en este ejemplo hipotético? En primer lugar, la entidad detectó una carencia: algunos proyectos podrían demandar un profesor de refuerzo, pero no cuentan con recursos suficientes para analizar individualmente cada caso.

En segundo lugar, a partir de la información de otros proyectos, para los cuales decidieron en su día si enviar un profesor de refuerzo o no, crean un modelo de *machine learning* capaz de aprender qué características son más relevantes a la hora de tomar la decisión y, por lo tanto, capaz de predecir por sí solo si, en otros casos donde se cuente con las mismas variables, debería enviarse un profesor extra de refuerzo o no. Ningún equipo de expertos de la entidad ha decidido qué características son las relevantes para tomar la decisión. Ha sido el algoritmo quien ha detectado los patrones por sí mismo, aprendiendo qué ha tenido en cuenta la entidad en el pasado a la hora de tomar la decisión.

Eso permite poder estimar en pocos segundos y para todos los proyectos si necesitan o no una persona de refuerzo, sin tener que llevar a cabo una evaluación individual y muy costosa de cada centro. La entidad gana en eficiencia y efectividad, pudiendo poner el foco en aquellos proyectos que más lo necesitan.

### 3. EL ORIGEN DE LA INTELIGENCIA ARTIFICIAL Y LOS ÚLTIMOS AVANCES

La historia del desarrollo de la inteligencia artificial es una historia de idas y venidas, de momentos álgidos y momentos de estancamiento. Desde modelos más básicos hasta los sofisticados sistemas de aprendizaje profundo. Un avance no lineal que ha «explotado» en el último lustro. Ahora bien, ¿cómo hemos llegado hasta aquí?

Los orígenes de la IA se remontan a mediados del siglo XX, cuando **en la década de 1950** Alan Turing propuso la idea de las «máquinas de pensar». En 1956 se celebró la Conferencia de Dartmouth, donde se acuñó el término *inteligencia artificial*. John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon, organizadores de la conferencia, son considerados algunos de los padres fundadores de la IA.

Uno de los primeros modelos de aprendizaje automático o *machine learning* fue el perceptrón. Un algoritmo de aprendizaje supervisado que servía para resolver tareas sencillas de clasificación. Pese a que inicialmente generó mucho entusiasmo, pronto se hicieron evidentes sus limitaciones a la hora de resolver problemas más complejos.

**En los años 60 y 70**, la IA experimentó lo que se conoce como su primera «época dorada», con avances en áreas como el procesamiento del lenguaje natural y los sistemas expertos. Durante esa época, los investigadores creían que estaban a las puertas de crear máquinas que podrían entender y responder al lenguaje natural, resolver problemas complejos y aprender de su experiencia. No obstante, esos sistemas no «aprendían» de la misma forma que lo hacen las modernas técnicas de *machine learning*, como las redes neuronales. Se basaban en reglas explícitamente codificadas por humanos. No aprendían de los datos ni mejoraban su rendimiento con la experiencia.

Esas limitaciones comportaron, de nuevo, que el entusiasmo inicial empezase a disiparse a **mediados de la década de los 70**, dando lugar al «invierno de la IA». Como resultado, se redujo la financiación para la investigación, y el campo se convirtió en un tema menos popular en el ámbito académico.

**La década de los 80** marcó un hito importante en el desarrollo de las redes neuronales y, por extensión, de la inteligencia artificial. Hasta ese momento, la aplicación de los algoritmos de redes neuronales se había visto limi-

tada por las dificultades para entrenarlas de forma efectiva, en particular para las redes de diversas capas o «profundas». No se había encontrado una solución eficiente a la hora de entrenar esos modelos.

La situación cambió cuando se produjo un avance crucial en la forma de entrenar estas redes: el algoritmo de propagación hacia atrás, o *backpropagation*, publicado por Rumelhart, Hinton y Williams en 1986. Este algoritmo abrió la puerta al entrenamiento eficaz de redes neuronales más profundas y complejas, lo que finalmente condujo al desarrollo de las técnicas de aprendizaje profundo que vemos hoy.

No obstante, aunque el algoritmo de *backpropagation* fue un avance importante, las limitaciones en la capacidad de computación de la época y la falta de grandes conjuntos de datos impidió la generalización de las redes neuronales. En otras palabras, en los años 80 ya existían algoritmos de IA con un gran potencial, pero escaseaban los buenos datos y ordenadores suficientemente potentes.

Esa limitación continuó durante **la década de los 90**, aunque eso no significó que no se produjesen avances importantes. El *machine learning*, y especialmente el aprendizaje supervisado, empezó a ganar más relevancia durante estos años. Se desarrollaron y refinaron una serie de algoritmos, como las máquinas de vectores de soporte (SVM), que proporcionaron un rendimiento sólido en una gran variedad de tareas.

Ahora bien, no fue hasta **la primera década de los años 2000**, con la llegada de la computación de alto rendimiento y el *big data*, cuando se crearon las condiciones adecuadas para que las redes neuronales, y por lo tanto el aprendizaje profundo, empezasen a mostrar su verdadero potencial. La popularización de internet y la proliferación de dispositivos digitales generaron una gran cantidad de datos, proporcionando el combustible necesario para los algoritmos de *machine learning* y *deep learning*.

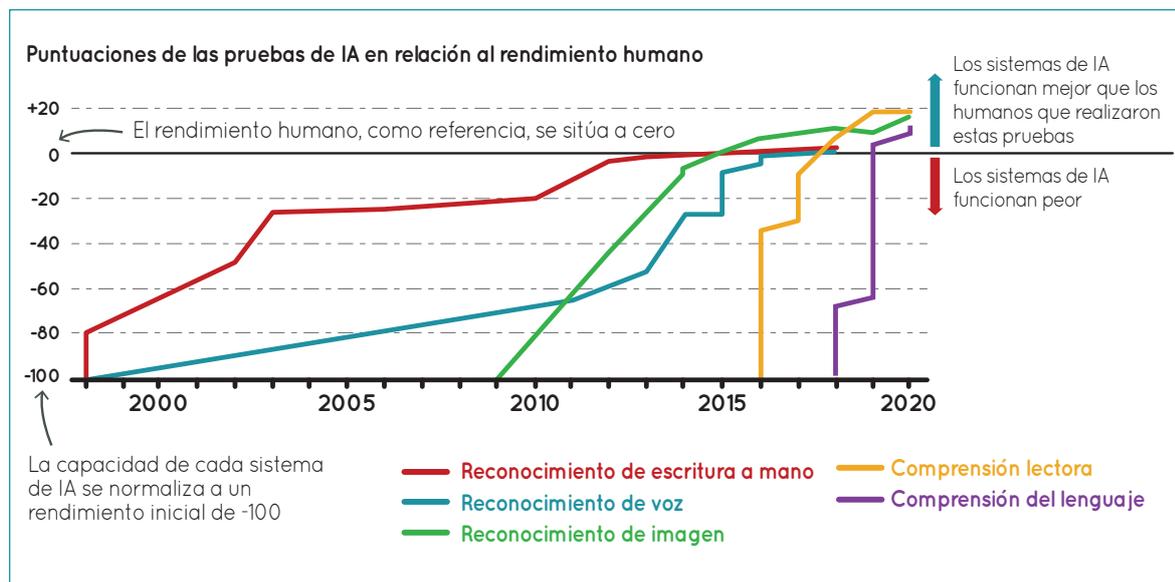
**En 2006, Geoffrey Hinton publicó un artículo que detallaba cómo entrenar redes neuronales profundas, lo que marcó el inicio de la era moderna del aprendizaje profundo. Desde entonces, los avances en la IA han sido rápidos y continuos.**

**La década del 2010** fue un periodo de gran avance en el campo de la inteligencia artificial. Las redes neuronales convolucionales (CNN), introducidas por Yann LeCun, han transformado el campo del reconocimiento de imágenes, convirtiéndose en la herramienta estándar para tareas de visión por computadora como el reconocimiento de objetos en imágenes. En el campo del procesamiento del lenguaje natural (NLP), las redes neuronales recurrentes (RNN) demostraron ser muy efectivas y permitieron avances

significativos. Además, en 2017 se produjo un avance que marcaría un antes y un después: Google Research publicó la arquitectura de red neuronal *Transformer*. Los *Transformers* se han utilizado para construir los modelos GPT y LaMDA, que están detrás de las interfaces de ChatGPT, de OpenAI, y Bard, de Google. Modelos capaces de producir respuestas y soluciones al mismo nivel o superior a un humano.

El siguiente gráfico probablemente sea el mejor resumen posible de la rápida evolución de la IA en los últimos años. Publicado por *Our World in Data* en el artículo «The brief history of artificial intelligence: The world has changed fast – what might be next?», muestra la puntuación máxima conseguida por un modelo de IA en diferentes tareas a lo largo del tiempo. Una puntuación igual a 0 equivale a la puntuación obtenida por un humano, de forma que una vez que un modelo de IA consiga un resultado superior a 0, implica que ha superado las capacidades de una persona.

### Gráfico 1. Las capacidades de reconocimiento de imágenes y lenguaje de los sistemas de IA han mejorado rápidamente.



Fuente: Extret de Roser (2022b)

En 2020, diferentes modelos de IA eran capaces de puntuar igual o mejor que un humano en tareas como el reconocimiento de voz, de escritura o de imágenes, la comprensión lectora y la comprensión del lenguaje. En estos últimos ámbitos, el avance es sorprendentemente rápido, superando a los humanos en pocos años.

Hace tiempo que la IA se utiliza para una miríada de actividades. Fija el precio del billete de avión, monitoriza nuestro comportamiento en los aeropuertos y asiste al piloto del avión durante el vuelo. Existen modelos de IA que determinan la concesión de un préstamo, de una ayuda o de un puesto de trabajo. También se utilizan para transcribir y traducir texto. Los asistentes virtuales como Alexa están presentes en miles de hogares. Las recomendaciones de series, películas y vídeos que nos hacen Netflix o YouTube se basan en IA llamadas sistemas de recomendación. **La IA no es ni mucho menos una tecnología de futuro, es una tecnología de presente que ya ha impactado en nuestras vidas y que hace tiempo que lo hace.**

Según el informe del Observatorio Nacional de Tecnología y Sociedad (ONTSI) sobre el uso de la inteligencia artificial de las empresas españolas, el 11,8 % de las empresas con más de 10 empleados hacían uso de esta tecnología en 2022 (un 50 % más que el año anterior), y un 4,6 % en el caso de las microempresas (un crecimiento aproximado de un 30 %). España, en 2021, se situaba en la media de la UE27, aunque lejos del nivel de adopción de los países líderes como Dinamarca (24 %), Portugal (17 %) o Finlandia (16 %).

Entre las empresas, destaca la adopción de la IA para automatizar flujos de trabajo o ayudar en la toma de decisiones (46,2 % de las empresas) y para identificar objetos o personas en función de imágenes (39,7 %). Más de un 30 % utilizan técnicas de procesamiento del lenguaje natural, más de un 20 % utilizan tecnologías para generar lenguaje escrito o hablado y en torno a un 30 % aplican técnicas de *machine learning* (o aprendizaje automático). Además, estos porcentajes no dejarán de crecer. En la estrategia España Digital 2026 del Gobierno se fija la meta de que para 2025 el 25 % de las empresas españolas hagan uso de inteligencia artificial y *big data*.

No obstante, **no fue hasta la aparición de la IA generativa, y en concreto de la famosa aplicación ChatGPT, que la IA traspasó la opinión pública.** ChatGPT alcanzó 1.000.000 de usuarios en solo 5 días. Alcanzar esa cifra le llevó dos meses y medio a Instagram, cinco meses a Spotify, diez meses a Facebook y tres años y medio a Twitter. Un éxito histórico.

La capacidad del modelo GPT y su fácil interacción con el usuario a través de la interfaz ChatGPT han sorprendido al mundo entero. Un modelo de IA capaz de resolver tareas para las cuales no fue entrenado. Realiza traducciones razonablemente buenas, sintetiza información, resuelve dudas, genera código en casi cualquier lenguaje de programación, y un largo etcétera. Además, la generación de texto no es lo único de lo que es capaz la

inteligencia artificial. La IA generativa incluye tanto la generación de texto como de imágenes, sonidos o vídeos. En el campo de la generación de imágenes destaca DALL-E, de OpenAI, capaz de generar y/o editar imágenes a partir de texto, o Stable Diffusion, que además es de código abierto.

A día de hoy, todos estos sistemas presentan ciertas limitaciones. Las imágenes generadas pueden contener errores. ChatGPT puede alucinar información<sup>4</sup> o errar en cálculos matemáticos. No obstante, no conviene perder la perspectiva temporal: esta es la primera generación de estos modelos, creados en tan solo unos pocos años. Es posible que, pese a tratarse de limitaciones complejas, puedan quedar resueltas con el tiempo. Probablemente será entonces cuando podremos evaluar el verdadero potencial de los sistemas de IA.

Así pues, a lo largo de las décadas, la IA ha pasado de ser un sueño a una realidad que toca casi cada aspecto de nuestras vidas. El reconocimiento de imágenes, la transcripción de audios, la traducción de textos, realizar tareas de predicción o la generación de texto o vídeos son ámbitos que han quedado dominados por modelos de inteligencia artificial.

Como no puede ser de otra forma, y como veremos en capítulos posteriores, la IA también puede aplicarse en el contexto de la Administración pública, ofreciendo oportunidades para mejorar la eficiencia, la precisión y el alcance de los servicios públicos. Pero también generando riesgos y desafíos. **Entender el contexto presente es fundamental para planificar las acciones de futuro en el ámbito del tercer sector.**

---

4 En el contexto de modelos de lenguaje, «alucinar información» hace referencia al hecho de que el modelo puede generar o proporcionar detalles que no están apoyados por datos reales o información verdadera que haya aprendido. Eso puede ser causado por la propia naturaleza del entrenamiento del modelo, dado que no está enfocado a «memorizar» datos ni hechos, sino que el modelo ve vastas cantidades de texto y aprende a predecir la palabra siguiente en una secuencia.

## 4. GOBERNANZA Y ESTRATEGIAS NACIONALES EN EL ÁMBITO DE LA INTELIGENCIA ARTIFICIAL

¿Cómo se toman las decisiones? ¿Están alineadas con valores éticos? ¿Quién las toma? ¿Cómo se implementan y quién las supervisa? En definitiva, ¿por qué es importante hablar de la gobernanza de la IA? Por un lado, la inteligencia artificial tiene el potencial para transformar nuestras sociedades. **Aplicar una gobernanza adecuada en el momento correcto puede contribuir a conseguir que la transformación sea positiva, previniendo y minimizando los riesgos y potenciando los beneficios.** Por otro lado, todo sistema de IA que esté siendo utilizado por una empresa o institución pública tendría que poder estar sometido a un marco universal que aprobase una adecuada utilización.

Por estos motivos, tras haber indagado en el desarrollo y evolución de la IA, es indispensable conocer cómo está siendo la interacción entre gobiernos, empresas, sociedad civil y ciudadanía y, por lo tanto, cómo se está configurando el ambiente regulador y discursivo en torno a la IA.

La gobernanza de la inteligencia artificial se ejecuta en diferentes niveles y ámbitos: puede desarrollarse a nivel internacional o nacional, puede provenir y estar dirigida exclusivamente al sector privado, al sector público o a ambos. Engloba aspectos como la definición de estándares éticos, la promulgación de leyes y regulaciones y la implementación de sistemas de supervisión y control para prevenir y mitigar posibles daños o abusos.

En este sentido, existen diferentes tipos de iniciativas de gobernanza global en el área de la IA. Algunas de estas iniciativas parten de las empresas privadas. Por ejemplo, en Microsoft han impulsado la Oficina de IA Responsable (ORA) y el Comité de AI y Ética en Ingeniería e Investigación (Aether) para garantizar internamente que los sistemas de IA se desarrollen de forma responsable. O la Partnership on AI, fundada en 2016 por Amazon, Facebook, Google, DeepMind, Microsoft, IBM y Apple (que entró en 2017), que agrupa a más de 100 socios del ámbito académico, la sociedad civil y la industria en una organización sin ánimo de lucro que aboga por la investigación en buenas prácticas y el uso responsable de la IA. No obstante, en los últimos tiempos se están produciendo cambios. Compañías como Microsoft, Twitter o Twitch están prescindiendo de los trabajadores que estudiaban el lado negativo de la IA.<sup>5</sup>

---

<sup>5</sup> Información extraída del artículo de *The Washington Post* «As AI booms, tech firms are laying off their ethicists», disponible en: <https://www.washingtonpost.com/technology/2023/03/30/tech-companies-cut-ai-ethics/>.

También pueden provenir de organismos internacionales, como las Recomendaciones en IA de la OCDE de 2019 o las Recomendaciones sobre la ética de la inteligencia artificial de la UNESCO de 2021. En ambos casos no dejan de ser normativas no vinculantes que fundamentalmente aluden a principios éticos y morales.

No obstante, en cuanto al poder, estamos en un sector con una concentración considerablemente elevada. Un reducido número de empresas controlan los recursos que facilitan la creación de los sistemas de IA, incluyendo los recursos físicos, epistémicos e informáticos. Además, la forma en que se distribuye la IA, normalmente en formato de plataformas, las convierte en unos actores importantes. Esta circunstancia proporciona un papel clave a las principales firmas, que pueden tratar de impulsar las expectativas, los marcos de debate y los discursos sobre el presente y el futuro de la IA, la visión, los intereses y las propuestas.

El principal riesgo de esta situación es que se implementen tecnologías que, por falta de visiones críticas, produzcan más riesgos que beneficios. **Por eso es fundamental que otros actores, como el tercer sector, se organicen, coordinen y fijen una postura conjunta.** Una postura a favor de los avances y mejoras que pueda traer la tecnología, pero previsoramente ante sus riesgos, y que ofrezca garantías en su aplicación.

Hasta ahora, dos gigantes tecnológicos han tomado claramente la delantera. Microsoft se ha aliado con OpenAI. Este laboratorio es el precursor de la herramienta más famosa, ChatGPT, pero también hay otros, como DALL-E 2 (capaz de generar imágenes a partir de unas pocas palabras) y Whisper (transcripción de audio en inglés a un nivel humano). Tras invertir 10.000 M€ en OpenAI, Microsoft está integrando la IA en todos sus productos: Windows, Word, etc. Por su parte, Google ha respondido aliándose con DeepMind, una empresa especializada en IA con sede en Londres. Google ha publicado la herramienta Bard, competidora directa de ChatGPT. Dos gigantes tecnológicos y dos laboratorios de investigación de IA que (por ahora) están liderando la competición.

Con todo, en los últimos meses Meta ha irrumpido con fuerza. Ha publicado modelos de generación de lenguaje, generación de imagen, generación de texto a voz, y viceversa, y generación de música. La gran diferencia de Meta respecto a sus competidores es que ha apostado por ser *open source* (de código abierto), de forma que cualquier programador del mundo puede acceder al modelo y al código, y puede contribuir a mejorarlo y desarrollarlo. Eso aumenta considerablemente la capacidad para crear modelos más eficientes, más robustos y más seguros.

## 4.1 El rol de los estados en la gobernanza de la IA

Ahora bien, ¿qué rol están tomando los estados? Uno de los cambios fundamentales que se están produciendo es el papel que están asumiendo los países. Los estados de la Unión Europea están liderando este nuevo rol, tanto por su impulso conjunto a la regulación de la IA como por las publicaciones de sus estrategias nacionales en este ámbito.

En un estudio reciente,<sup>6</sup> los países eran clasificados en función del enfoque y acciones por las que se decantaban en sus estrategias nacionales:

- ) Los **países desarrolladores**: priorizan un rol del Estado como facilitador e iniciador de los proyectos de IA a través de la innovación y la asignación directa de recursos. Este enfoque sería el predominante entre los países del bloque postsoviético y de Asia del Este, donde el Estado asumiría el rol principal en el desarrollo.
- ) Los **países que asumen un rol de promoción**: otorgan el liderazgo al sector privado. Países como Estados Unidos, Reino Unido o Irlanda toman un enfoque descentralizado y delegan el rol clave en el sector privado, no implicando al Estado y apostando por la autorregulación de la industria. Aunque con enfoques diferentes, estos países priorizarían la innovación por delante de la protección.
- ) Los **países que apuestan por el control**: los países de la Unión Europea sitúan el centro de atención sobre el control, reflejando la fuerte posición de la Unión en cuanto a la regulación. Los estados quieren garantizar la protección de la sociedad de los riesgos de la IA mediante rigurosos marcos reguladores, priorizando la protección por delante de la innovación.

**El rol del Estado es relevante porque abre o cierra ventanas de oportunidad para actores como las organizaciones del tercer sector.** En un contexto donde el Gobierno delegue el liderazgo en el sector privado, será más difícil para las entidades que su visión y perspectiva sean influyentes. No obstante, este no es el contexto europeo, lo que **abre potencialmente la puerta a las entidades tanto para hacer llegar sus preocupaciones como para hacer contribuciones significativas.** No solo en términos legislativos, sino también de discurso.

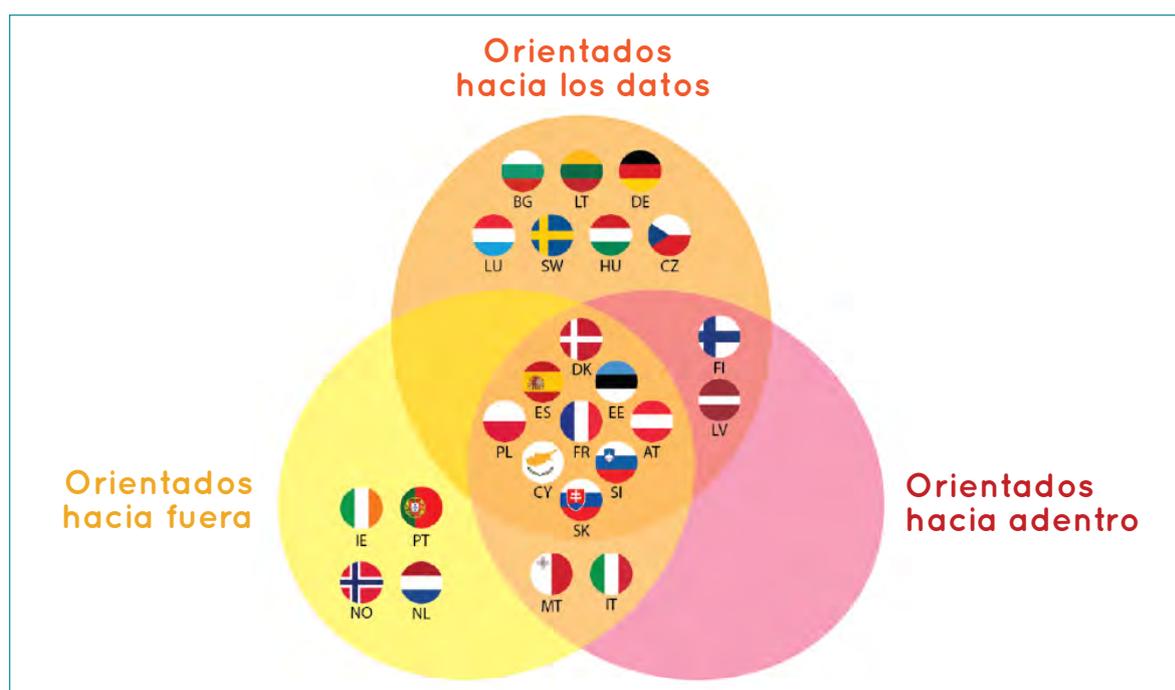
6 Gleb Papyshv & Masaru Yarime (2023) The state's role in governing artificial intelligence: development, control, and promotion through national strategies. *Policy Design and Practice*, 6:1, 79-102, DOI:10.1080/25741292.2022.2162252.

## 4.2 El rol de los estados en la aplicación de la IA en el sector público

El informe *European Landscape on the Use of Artificial Intelligence by the Public Sector*, elaborado por el servicio de supervisión de desarrollo e impacto de la IA de la Comisión Europea (*AI Watch*), analiza 24 estrategias nacionales de los países de la UE (más Noruega) y las clasifica en tres tipos ideales en función del grado de implicación, acciones e iniciativas del sector público. Este tipo de ejercicios son muy útiles, ya que permiten posicionar a los gobiernos de forma relativa a las acciones del resto. Saber qué papel está adoptando nuestro país es necesario para comprender qué papel puede jugar el tercer sector en la gobernanza de la IA.

La primera tipología de estados son aquellos **orientados hacia fuera (externally-oriented)**. Estos gobiernos ponen el foco en la cooperación público-privada para desarrollar la IA en el sector público. Asumen que el Estado no tiene suficiente capacidad y competencias y que los sistemas son demasiado complejos como para enfrentarse a la IA de forma autónoma. Por ese motivo, ponen mayor énfasis en crear un ecosistema de *start-ups* y empresas que desarrollen IA para el sector público, promoviendo la colaboración público-privada. En este grupo se encuentran países como Irlanda, Portugal, Noruega y Países Bajos.

**Gráfico 2.** Ejercicio de agrupación de las estrategias nacionales.



Fuente: extraído de Tangi *et al.* (2022)

Por otro lado, los **países orientados hacia los datos (data-oriented)** priorizan facilitar el acceso y mejorar la calidad de los datos, eliminando las barreras y mejorando las infraestructuras para el desarrollo de la IA en el sector público. Tornar disponibles más bases de datos, fomentar la compartición de datos entre instituciones públicas o mejorar las prácticas de recogida y gobierno de los datos serían sus acciones principales. Entre los países que encajan en este tipo ideal están Alemania o Suecia.

Los estados con una estrategia **orientada hacia dentro (internally-oriented)** son aquellos que se centran en mejorar la capacidad interna del Estado como instrumento para estimular el desarrollo de la IA. Crear nuevas unidades o instituciones públicas especializadas en IA, mejorar el capital humano del funcionariado mediante formación y aumentar el conocimiento interno sobre IA mediante eventos serían sus acciones predilectas. No obstante, ningún país apuesta exclusivamente por estas medidas como solución única. Por ejemplo, Finlandia y Letonia lo combinan con una estrategia orientada hacia los datos.

Finalmente, hasta 9 de los estados analizados, entre los cuales se encuentra España, integran los 3 enfoques en sus estrategias nacionales. Por lo tanto, **existe una clara apuesta en la mayoría de países por trabajar conjuntamente con el sector privado en el impulso de la IA en el sector público, lo que es razonable, ya que es donde está la mayor parte de capital humano y tecnológico en la actualidad.**

### 4.3 La estrategia nacional de España

En el caso particular de España, el Gobierno estatal publicó su Estrategia Nacional de Inteligencia Artificial en diciembre de 2020, con el objetivo de desarrollar un marco político que defina las diversas acciones que las administraciones gubernamentales emprenderán para facilitar el desarrollo y despliegue de la IA en la economía y la sociedad.

En la estrategia nacional destacan objetivos vinculados a la formación y la investigación, como es el impulso al desarrollo del capital humano en IA mediante formación, atracción de talento y empleo cualificado, junto con el desarrollo de una excelencia científica para promover España como país líder.

También se fija como objetivo impulsar el despliegue y uso de tecnologías de IA tanto en el sector público como en el privado, y situar España como líder en el desarrollo de herramientas, tecnologías y aplicaciones para la proyección y uso del idioma español en IA.

A la vez que se promueve el desarrollo del capital humano y tecnológico, también se fijan objetivos éticos, como garantizar un marco ético que delinee los derechos individuales y colectivos y construya un entorno de confianza, así como garantizar la inclusión en la economía impulsada por la IA –inclusión de género, digital o territorial.

En cuanto a medidas directamente relacionadas y con un impacto en el sector público, la estrategia nacional de España presenta iniciativas políticas en todas las áreas analizadas por AI Watch:

- (1) Aumentar el conocimiento de los empleados públicos sobre IA.
- (2) Mejorar la capacidad interna, formando a las funciones en IA y creando nuevos departamentos especializados.
- (3) Medidas para mejorar la calidad y el acceso a los datos.
- (4) En cuanto a la ética y el marco legal, se quiere desarrollar tanto un marco ético como reformar las leyes para el intercambio de datos.
- (5) En cuanto a la financiación y los procesos burocráticos, existe un compromiso para financiar proyectos de IA, estimular las *start-ups* del *GovTech* y revisar los procesos burocráticos para agilizarlos.
- (6) En el área del *learning by doing*, España quiere regular y liderar la implementación de los *sandboxes*.

Por lo tanto, independientemente de la traducción definitiva en iniciativas políticas y marcos normativos, y siempre que no cambien las prioridades o los partidos de gobierno, **existe a priori una clara apuesta para que España sea líder en el desarrollo y aplicación de la inteligencia artificial, tanto en el sector privado como público.**

## 5. EL ENFOQUE CENTRADO EN LA PERSONA Y LA LEY DE INTELIGENCIA ARTIFICIAL DE LA UNIÓN EUROPEA

En el capítulo 3 se ha expuesto el desarrollo histórico y la evolución de la inteligencia artificial desde mediados del siglo XX hasta hoy. Después de indagar en esta perspectiva, el foco del capítulo 4 se ha puesto en discernir cómo los gobiernos, las empresas y la sociedad civil están interactuando entre sí, asentando las bases de la gobernanza de la IA del futuro, y, particularmente, qué aplicación de la IA en las administraciones públicas están siguiendo los estados.

El desarrollo de cualquier sector económico está fuertemente determinado por el marco regulador en el que opera. En el ámbito de la inteligencia artificial, la ley que más debate está generando por su potencial para impulsar (o socavar) el desarrollo de la IA en los países europeos es la Ley europea de Inteligencia Artificial.

Esta ley parte de una visión muy concreta: el enfoque centrado en la persona. Una visión que aboga por un progreso «útil» para la persona. Es decir, entiende la IA como un medio para mejorar el bienestar, y no como un fin en sí mismo. En este apartado se expone este enfoque para, una vez comprendido el punto de partida, explicar las claves del borrador de ley que ya ha sido aprobado.

### 5.1 El enfoque centrado en la persona

En 2018, la Comisión Europea creó el Grupo de Expertos de Alto Nivel en Inteligencia Artificial con el objetivo de desarrollar dos informes: (1) una guía ética sobre la IA (AI Ethics Guidelines) y (2) recomendaciones de políticas públicas e inversión.<sup>7</sup>

El grupo publicó un primer borrador de la guía ética en diciembre de 2018, que fue sometido a consulta con los *stakeholders* y los Estados miembro. En abril de 2019 se publicó el documento definitivo: Directrices éticas para una IA fiable (Ethics Guidelines For Trustworthy AI). Este documento, que fija también la posición de la Comisión Europea, transmite un mensaje claro: la IA no es un fin en sí misma, sino una herramienta que tiene que

---

<sup>7</sup> Los informes del grupo de expertos pueden consultarse en: <https://digital-strategy.ec.europa.eu/es/policies/expert-group-ai>.

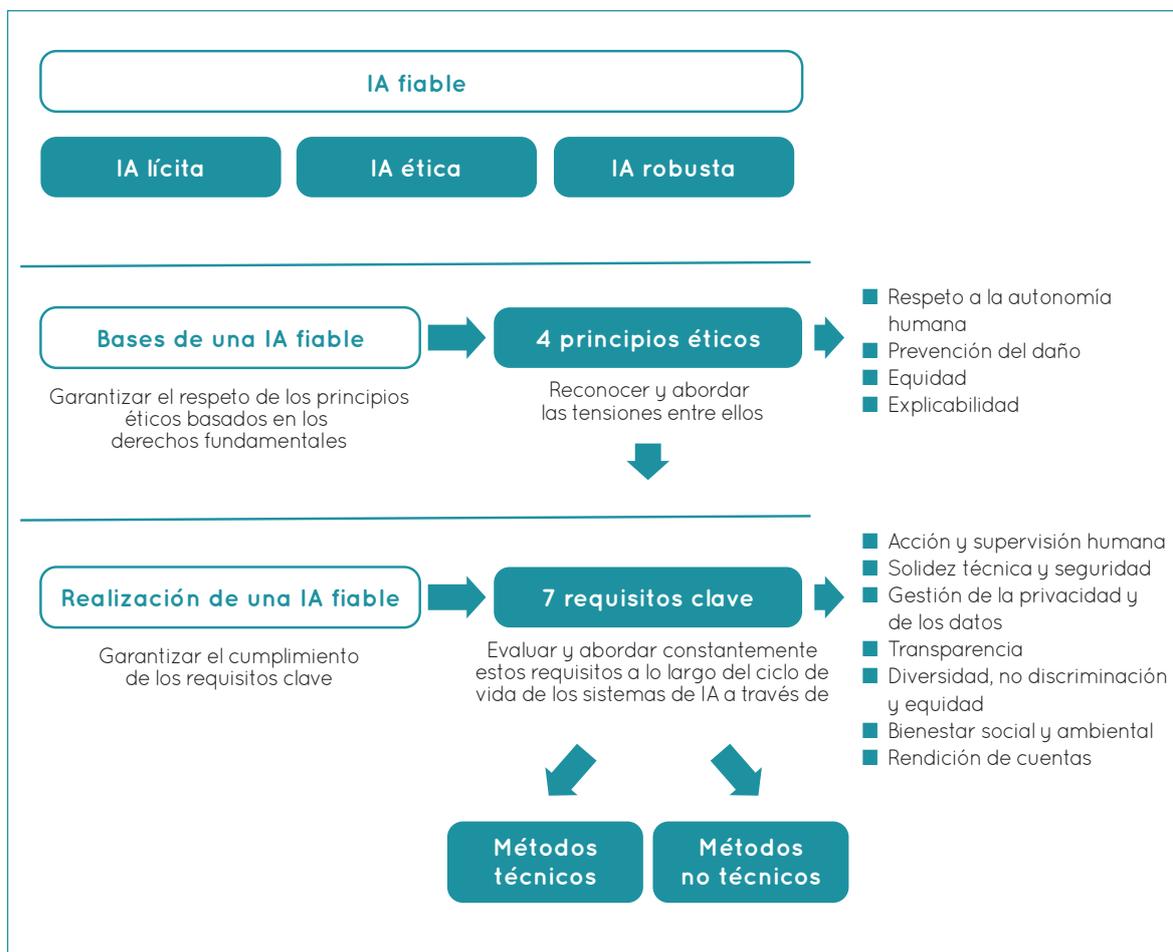
servir a las personas con el objetivo último de mejorar su bienestar. En otras palabras, **el desarrollo e implementación de la IA deben realizarse poniendo a las personas en el centro. Para conseguirlo, la confianza en la IA es un prerequisite.**

La IA tiene el potencial de transformar significativamente la sociedad, mejorando el bienestar individual y social. Pero, para que eso ocurra, los sistemas de IA deben estar centrados en la persona y deben basarse en un compromiso con su uso al servicio de la humanidad y el bien común. El momento presente es clave porque nos encontramos ante una importante ventana de oportunidad para dar forma al futuro (y presente) desarrollo de la inteligencia artificial, lo que implica tanto tratar de maximizar sus beneficios como prevenir y minimizar sus riesgos.

Ahora bien, ¿cómo puede garantizarse el desarrollo de una IA centrada en las personas? La propuesta del grupo de trabajo empieza por establecer qué derechos fundamentales y principios éticos debería respetar cualquier sistema que utilice inteligencia artificial. Entre ellos estaría el respeto a la libertad individual, la democracia, la igualdad o el Estado de derecho. A partir de todos los derechos y principios identificados, el grupo de trabajo resuelve que una IA es fiable si cumple tres componentes esenciales:

- ) **Lícita** (*lawful*): cumple todas las leyes y regulaciones aplicables.
- ) **Ética** (*ethical*): se adhiere a principios y valores éticos.
- ) **Robusta** (*robust*): es robusta tanto desde una perspectiva técnica como social

Una IA lícita, ética y robusta será una IA fiable, que respete los derechos fundamentales y los principios éticos citados. Estos 3 componentes abstractos se aterrizan en 7 requisitos que todo sistema de IA debería cumplir, que se resumen en el siguiente gráfico y se describen a continuación:

**Gráfico 3.** Marco para una IA fiable

Fuente: extraído y adaptado de HLEGAL (2019)

- (1) Agencia y supervisión humanas: el bienestar del usuario debe ser central para la funcionalidad del sistema, y la supervisión humana debe estar asegurada.
- (2) Solidez técnica y seguridad: una IA fiable requiere algoritmos seguros, fiables y robustos que hagan frente a los errores e inconsistencias que puedan surgir, así como contra posibles ataques.
- (3) Privacidad y gobernanza de datos: el respeto de la privacidad, la calidad y la integridad de los datos debe estar garantizado.
- (4) Transparencia: incluye la trazabilidad, la explicabilidad y la comunicación.
- (5) Diversidad, no discriminación y equidad: incluye la prevención de prejuicios injustos, la accesibilidad y el diseño universal, y la participación de las partes interesadas.

- (6) Bienestar social y ambiental: los impactos deben ser considerados en términos sociales y no solo individuales, por lo que deben tenerse en cuenta tanto sus impactos sociales como medioambientales.
- (7) Rendición de cuentas: deben existir mecanismos que aseguren la rendición de cuentas de los resultados, antes y después de su implementación. Las auditorías (externas e internas) son fundamentales en este aspecto. También deben identificarse los potenciales impactos negativos, así como garantizar las compensaciones y reparaciones.

Por lo tanto, la visión de la Comisión Europea puede resumirse en la siguiente idea: **dada la capacidad transformadora de los sistemas de inteligencia artificial, nos encontramos ante una ventana de oportunidad para definir cómo debe desarrollarse e implementarse, de forma que nos aseguremos de que los sistemas de IA futuros situarán siempre a las personas en el centro, orientando los avances hacia la mejora del bienestar individual y social, y previniendo desde ahora mismo sus potenciales usos y diseños con consecuencias negativas.** Bajo esta visión se está desarrollando la Ley de Inteligencia Artificial de la Unión Europea.

## 5.2 La Ley de IA de la Unión Europea

En junio de 2023 se aprobó en el Parlamento Europeo el borrador de la Ley de Inteligencia Artificial de la Unión Europea (EU IA Act), una ley que se había estado preparando durante dos años. Ahora han empezado las negociaciones entre los Estados, durante las cuales España podría ejercer un papel clave bajo la presidencia del Consejo de la Unión Europea, y se espera que a finales de año se publique el texto definitivo.

Este texto legislativo no es relevante únicamente por el impacto directo que podría tener sobre nuestro país, sino porque es un caso de referencia a nivel mundial. La regulación que está desarrollando la Unión Europea (y también China) servirá al resto de países para tomar nota tanto de los aciertos como de los errores.

**La regulación de la IA ha sido una cuestión política central en la Unión en los últimos años, y este borrador es el resultado de ese esfuerzo.** Después de la regulación «blanda» no vinculante de 2019 a través de las Directrices éticas para una IA fiable, la Comisión pasó en 2021 hacia un enfoque legislativo haciendo un llamamiento a la adopción de un nuevo marco regulador. Partiendo de un enfoque que sitúa a las personas en el centro, es decir, de garantizar el desarrollo de una IA fiable, se busca crear una regulación que cumpla dos objetivos paralelos: (1) promover el desarrollo de la IA y (2) gestionar sus riesgos asociados.

En otras palabras, esta ley propone regular los usos de la IA y no la IA en sí misma, buscando conseguir que la regulación no quede obsoleta con los avances de la tecnología. Ahora bien, antes de entrar en esta cuestión, ¿cómo se define la inteligencia artificial? En el punto 1), artículo 3, se define como «sistema de inteligencia artificial»:

*«El software que se desarrolla utilizando una o diversas de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyen en los entornos con los que interactúa.»*

Estos sistemas de IA especificados en el anexo I son:

- a) Estrategias de aprendizaje supervisado, incluido el aprendizaje supervisado, el no supervisado y el realizado por refuerzo.
- b) Estrategias basadas en la lógica y el conocimiento, especialmente la representación del conocimiento, la programación (lógica) inductiva, las bases de conocimiento, los motores de inferencias y deducción, los sistemas expertos y de razonamiento (simbólico).
- c) Estrategias estadísticas, estimación bayesiana, métodos de búsqueda y optimización.

Por lo tanto, los productos comerciales de empresas privadas y aplicaciones de IA en la Administración pública que hagan uso de algunas de estas estrategias quedarán regulados (en principio) por esta ley.

Los objetivos específicos del reglamento son:

- 1) Garantizar que los sistemas de IA introducidos y utilizados en el mercado de la UE sean seguros y respeten la legislación vigente en materia de derechos fundamentales.
- 2) Garantizar la seguridad jurídica para facilitar la inversión e innovación en IA.
- 3) Mejorar la gobernanza y la aplicación efectiva de la legislación vigente en materia de derechos fundamentales y los requisitos de seguridad aplicables a los sistemas de IA.
- 4) Facilitar el desarrollo de un mercado único para hacer un uso legal, seguro y fiable de las aplicaciones de IA y evitar la fragmentación del mercado.

Como se ha comentado, esta legislación se basa en una clasificación de los riesgos que la inteligencia artificial podría suponer para la seguridad, la salud o los derechos fundamentales de una persona en cada caso, y de-

termina obligaciones para proveedores y usuarios en función de los riesgos que supongan los usos de la IA.

#### Gráfico 4.



Fuente: Ley de Inteligencia Artificial de la UE.

En primer lugar, los **sistemas de IA de riesgo bajo o mínimo** no estarán sujetos a ninguna obligación, y podrán ser desarrollados y utilizados en la UE sin más requisitos legales. No obstante, la ley propone la creación de códigos de conducta.

En segundo lugar, estarían los **sistemas de IA de riesgo limitado**. Sus responsables estarían sometidos a unas obligaciones generales de transparencia. Por ejemplo, deberán presentar información para comunicar el uso de un sistema de IA cuando este interactúe con humanos, o en caso de que se utilice para generar o manipular imágenes, audios o vídeos, será obligatorio informar de que se han generado por medios automatizados.

En tercer lugar, se encuentran los **sistemas de IA de alto riesgo**. En estos casos, el borrador cita explícitamente que «los requisitos relativos a los datos de alta calidad, la documentación y la trazabilidad, la transparencia, la vigilancia humana, la precisión y la solidez son estrictamente necesarios para reducir los riesgos de la IA para los derechos fundamentales y la seguridad».

Los sistemas considerados de alto riesgo se especifican en el punto 1 del artículo 6 del título III y en el anexo III. Sin enumerar todos los casos posibles, algunos ejemplos son:

- ) Sistemas utilizados como componentes de seguridad de un producto o como un producto sujeto a la legislación armonizada de salud y seguridad de la Unión (juguetes, aviación, coches, etc.).
- ) La identificación biométrica y categorización de las personas en tiempo real o de forma retroactiva.
- ) Sistemas utilizados para la gestión y operación de infraestructura crítica (tráfico de carreteras, suministro de agua, gas, calefacción y electricidad).
- ) Sistemas utilizados para evaluar estudiantes y determinar su acceso a instituciones educativas.
- ) Sistemas utilizados para seleccionar o descartar trabajadores o evaluar candidatos.
- ) Sistemas utilizados para evaluar el riesgo de incidencia criminal de una persona.
- ) Sistemas utilizados por las autoridades públicas para verificar la autenticidad de documentos de viaje.

Todos estos sistemas se estructuran en ocho áreas, que podrían ser actualizadas en caso necesario: identificación biométrica y categorización de personas; gestión y operación de infraestructura crítica; educación y formación; empleo y gestión del trabajo; acceso a servicios privados esenciales y servicios públicos; cumplimiento de la ley; migración, asilo y fronteras, y administración de justicia.

Estos sistemas estarán sometidos a toda una serie de nuevas reglas. Los proveedores de estos sistemas deberán registrarlos en una base de datos de la UE gestionada por la Comisión antes de comercializarlos para ser sometidos a una evaluación de conformidad ex ante. Además, deberán cumplir una serie de requisitos y obligaciones, en particular, en materia de gestión de riesgos, pruebas, solidez técnica, formación y transparencia, supervisión humana y ciberseguridad (artículos 8 a 15).

Finalmente, están los **sistemas de IA que suponen un riesgo inaceptable**. Estos sistemas están directamente prohibidos, ya que su uso se considera inaceptable por ser contrario a los valores de la Unión. Quedan prohibidos los siguientes usos:

- 1) El uso de técnicas subliminales que trasciendan la conciencia de una persona para alterar su comportamiento o provocar perjuicios físicos o psicológicos.
- 2) La IA que aproveche alguna de las vulnerabilidades de un grupo específico de personas debido a su edad o discapacidad física o mental para alterar su comportamiento.

- 3) La utilización de sistemas de IA por parte de las autoridades públicas con la finalidad de evaluar o clasificar la fiabilidad de personas físicas.
- 4) El uso de sistemas de identificación biométrica remota «en tiempo real» en espacios de acceso público (excepto en casos concretos como la búsqueda de víctimas de un delito o prevención de una amenaza).

Por otro lado, como medida de apoyo a la innovación, en el artículo 53 se recoge el establecimiento de un entorno controlado de pruebas de IA (*sandbox*), que será establecido por uno o más Estados miembro o por el supervisor europeo de protección de datos. Dicho entorno facilitará el desarrollo, prueba y validación de los sistemas de IA antes de ser introducidos en el mercado. En esta cuestión, España quiere posicionarse como país de referencia, tal y como se verá en el próximo apartado.

Finalmente, la ley propone la creación de un Comité Europeo de Inteligencia Artificial, que actuaría como consejo asesor de la Comisión con el objetivo de contribuir a la cooperación efectiva entre las autoridades nacionales y la Comisión, coordinar y contribuir a los análisis sobre cuestiones del mercado interno y asistir a las autoridades nacionales y a la Comisión sobre la aplicación de esta regulación. Dicho comité estaría compuesto por representantes de las autoridades nacionales. Estas autoridades deberán ser establecidas o designadas por cada uno de los Estados.

### 5.3 Las acciones de España

Hasta ahora, además de los documentos políticos como la Estrategia Nacional de Inteligencia Artificial, España ha aprobado dos decretos de especial interés para visualizar los pasos que está dando el país en este ámbito. En verano de 2023 se aprobaron dos decretos relevantes: por un lado, uno que regula la aplicación del *sandbox* en España y, por otro, los estatutos de la agencia supervisora.

#### Real Decreto de *sandbox* de España

En junio de 2022, el Gobierno de España y la Comisión Europea presentaron el proyecto piloto para poner en marcha el primer *sandbox* regulador de la UE sobre IA, que será financiado mediante los fondos de recuperación y resiliencia con 4,3 millones de euros hasta 2025.

Un año después, en junio de 2023, el Gobierno publicó el borrador del Real Decreto regulador del *sandbox* de inteligencia artificial. El objetivo es proporcionar a las empresas, especialmente a las pequeñas y medianas

empresas y las *start-ups*, un entorno controlado de pruebas que garantice el desarrollo de una tecnología responsable y centrada en la persona, mitigando los riesgos potenciales para la salud, la seguridad y los derechos fundamentales.

La participación en este entorno controlado de pruebas se circunscribe a sistemas de IA que sean clasificados como de alto riesgo, a sistemas de IA de propósito general, modelos fundacionales y sistemas de IA generativa. Por otro lado, se excluye la participación de sistemas con finalidad científica, de defensa o de seguridad nacional.

El objetivo final del sandbox es **que las empresas que estén desarrollando este tipo de aplicaciones tengan un entorno donde probarlas y evaluarlas, de forma que se asegure que se están cumpliendo todos los requisitos que exigirá el futuro Reglamento europeo de la IA de cara a poder comercializar su producto.**

### Estatuto de la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA)

En agosto de 2023 se aprueba el real decreto que da el visto bueno al estatuto de la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA), adscrita al Ministerio de Asuntos Económicos. Con su creación, España se ha convertido en el primer país europeo que cuenta con una entidad de estas características. Esta agencia será la encargada de supervisar la aplicación de la normativa comunitaria en materia de inteligencia artificial.

## 6. LA INTELIGENCIA ARTIFICIAL EN LAS ADMINISTRACIONES PÚBLICAS

### 6.1 ¿Qué entendemos por IA en el contexto de las AA. PP.?

Los apartados previos, donde se ha explicado qué es la inteligencia artificial y analizado su evolución, gobernanza y el borrador de la Ley de Inteligencia Artificial de la Unión Europea, han servido para entender de qué estamos hablando cuando nos referimos a inteligencia artificial, situar el debate y contextualizar el momento político y social en el que nos encontramos actualmente.

Llegados a este punto, el siguiente paso es poner el foco en un aspecto de especial interés para el tercer sector: ¿cómo se está aplicando la inteligencia artificial en las administraciones públicas? En otras palabras, ¿en qué procesos y servicios públicos se está incluyendo la IA con el objetivo de mejorar su efectividad? En este apartado se explicarán, a través de casos reales, cómo se está utilizando la IA desde el sector público y, en particular, en el ámbito de los derechos sociales.

Como hemos comentado, la inteligencia artificial también ha llegado a las administraciones públicas. No obstante, es importante diferenciar el proceso de digitalización que está experimentando el sector público de la propia introducción de inteligencia artificial en este ámbito. La transformación digital de las administraciones públicas comprende un conjunto de mecanismos tecnológicos que pueden, o no, incluir inteligencia artificial.

En este sentido, cabe mencionar que la principal tecnología utilizada por el sector público son los sistemas de decisión automatizada (o ADMS, por sus siglas en inglés), definidos como «unos procesos pensados para optimizar la toma de una decisión mediante el uso de algoritmos y grandes cantidades de datos».<sup>8</sup> Estos sistemas pueden utilizar IA (concepto ya definido anteriormente), pero existen otros que no.

Finalmente, otro concepto vinculado a los ADMS es el de caja negra, que se refiere a cuando los procedimientos y variables que utilizan los algoritmos de los ADMS no son fáciles de observar y, por lo tanto, son poco transparentes para las personas.

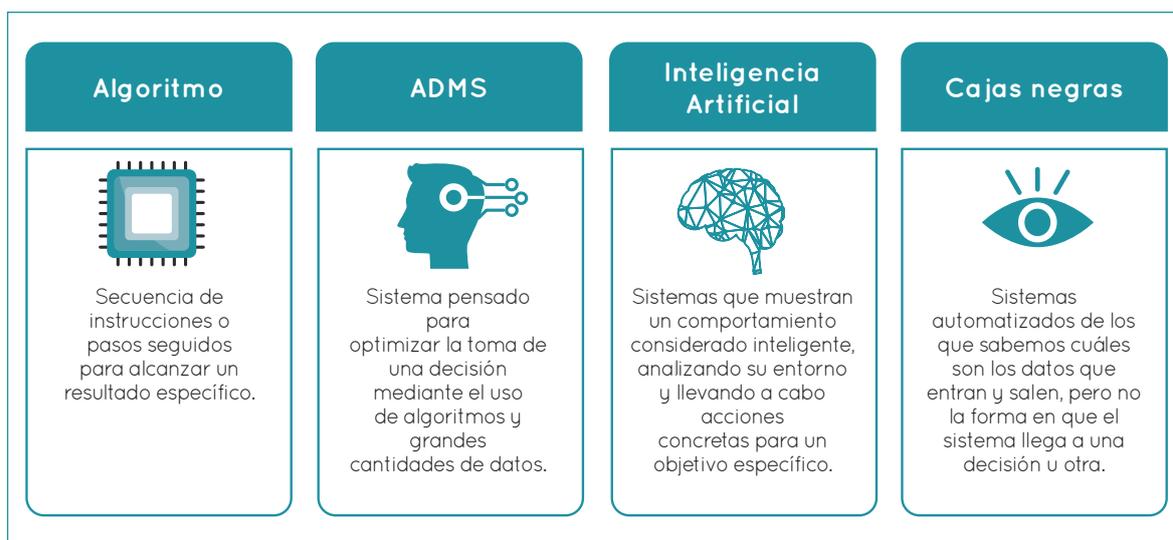
En el presente informe, pues, y tal y como se ha avanzado en la introducción, se utiliza el concepto de IA como un concepto amplio que incluye

---

<sup>8</sup> *AlgorithmWatch*, 2019, citado por Jiménez Arandia, P., 2023.

tanto los ADMS y la inteligencia artificial, dado que actualmente los ADMS se encuentran con mayor frecuencia dentro del sector público, y presentan oportunidades y riesgos similares a los de la IA.

### Gráfico 5.



Fuente: elaboración propia a partir de Jiménez Arandia, P. (2023).

## 6.2 Aplicaciones de la IA en las AA. PP.

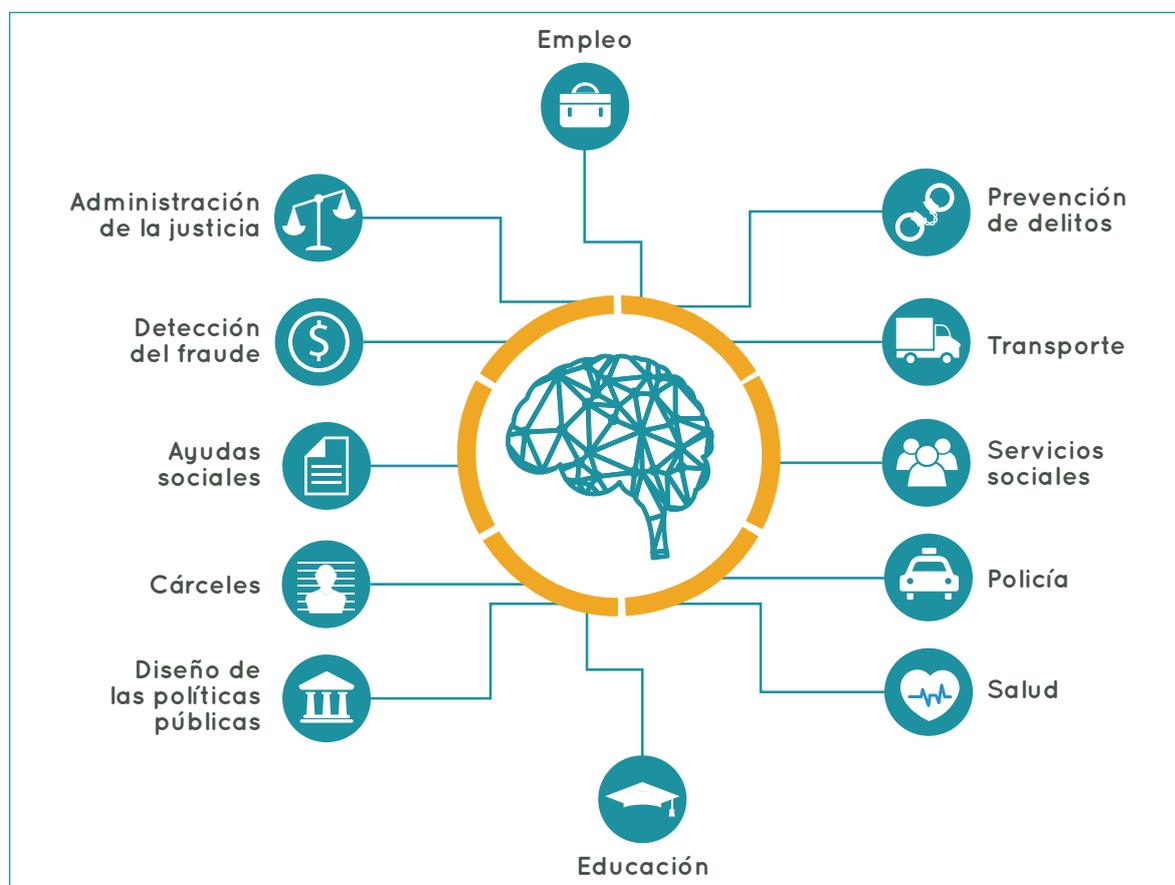
El uso de los ADMS y la IA en las administraciones públicas ha aumentado durante los últimos años. Como se ha mencionado anteriormente, el Gobierno se vincula con la IA a través de dos líneas de acción: por un lado, trabaja para establecer un marco legislativo que guíe su desarrollo y, por otro, trabaja para definir en qué tareas su implementación puede ser útil en términos de eficiencia y efectividad. En este apartado, pues, se expondrá en qué funciones dentro de la Administración pública se ha empezado a trabajar con IA.

Las experiencias internacionales sobre las primeras aplicaciones de IA dentro del sector público muestran cómo esta puede tener un gran impacto en la mejora de las operaciones internas, de la toma de decisiones, de los servicios públicos, así como de la confianza en el Gobierno. Esta evidencia, por lo tanto, señala el gran potencial que tiene la IA dentro del marco de las administraciones públicas.

**La inteligencia artificial permite, siempre que se mantengan unos principios éticos y propios de un estado democrático, que los procesos actuales realizados por la Administración pública sean más eficientes y precisos.** Tal y como se desarrollará con mayor detalle en la sección 7, los numerosos beneficios de la IA en el sector público pueden ayudar a diseñar mejores políticas públicas, mejorar la comunicación y el compromiso con los ciudadanos y mejorar la

calidad y velocidad de los servicios públicos. Por ese motivo, esta ha empezado a implementarse para automatizar las ayudas públicas (BOSCO), para detectar posibles casos de fraude al Estado (SyRI, Robodebt), para ayudar a la policía (VioGén, VeriPol, PredPol), para la prevención de delitos (RisCanvi, COMPAS, LSI-R, OASys), para fomentar el empleo, para acceder a la universidad (Ofqual) y para la detección de menores vulnerables desde los servicios sociales (Gladsaxe, Allegheny), entre otras funciones.<sup>9</sup>

## Gráfico 6.



Fuente: elaboración propia

Más concretamente, de acuerdo con la información recogida en el *Joint Research Centre* de la Comisión Europea,<sup>10</sup> las principales funciones en las que se ha empezado a trabajar con inteligencia artificial son en el ámbito del transporte; en órganos ejecutivos y legislativos, asuntos financieros y fiscales y asuntos exteriores; en servicios públicos generales; en servicios

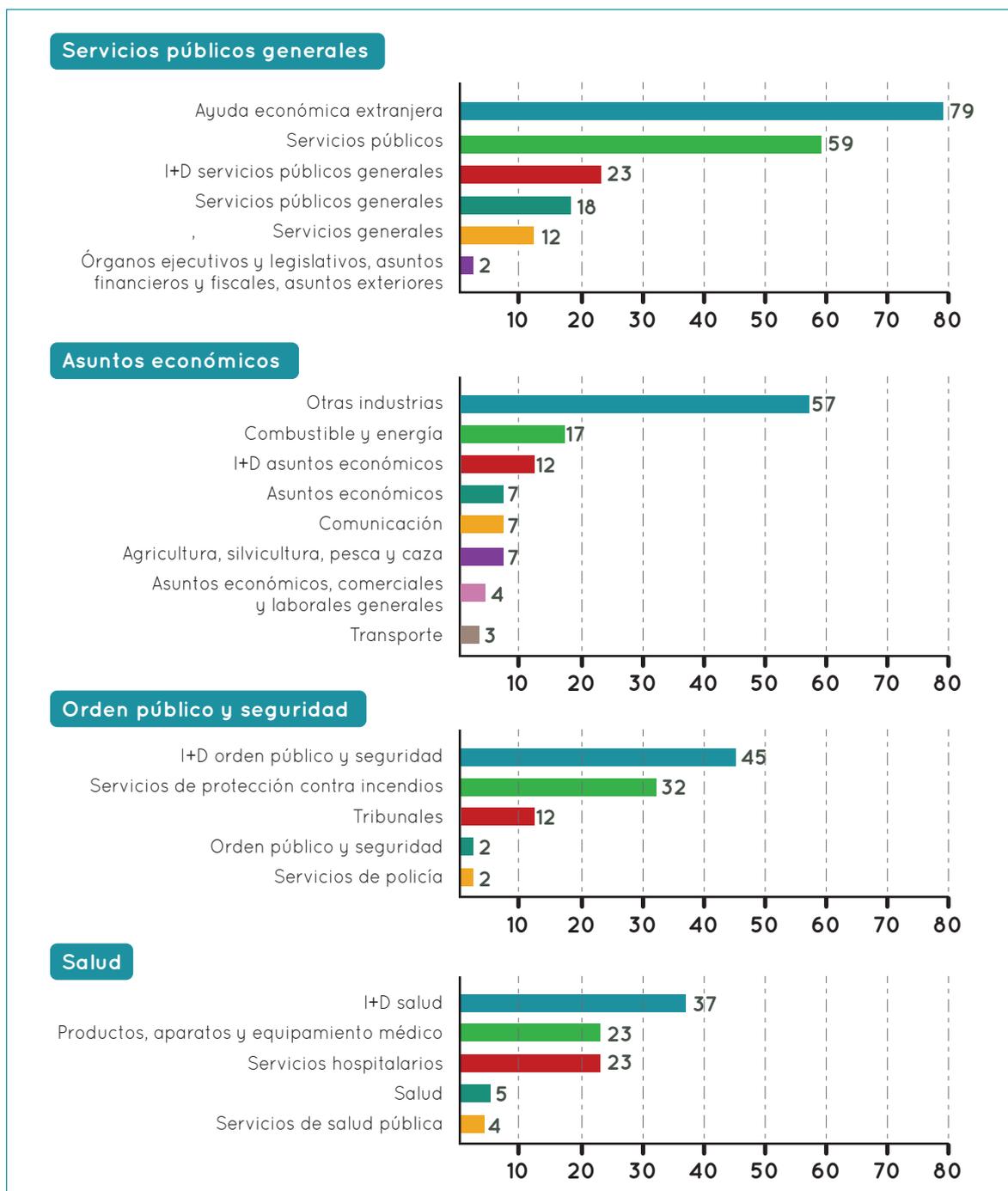
<sup>9</sup> A continuación se detallan los países que han implementado cada uno de los algoritmos mencionados:

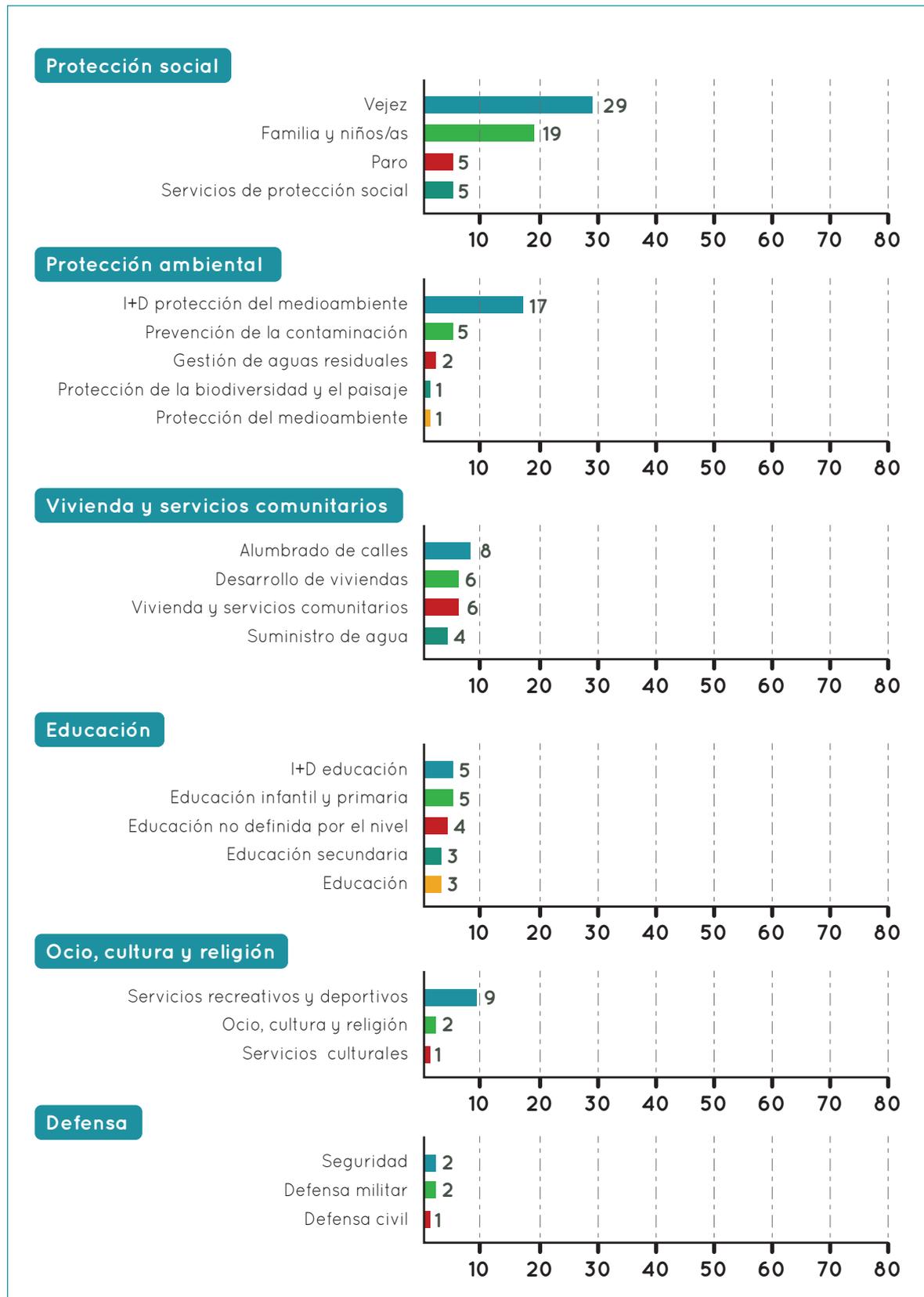
España: BOSCO, VioGén, VeriPol, RisCanvi; Holanda: SyRI; Australia: Robodebt; Estados Unidos de América: PredPol, COMPAS, LSI-R, Allegheny; Canadá: LSI-R; Reino Unido: LSI-R, OASys, Ofqual; Dinamarca: Gladsaxe.

<sup>10</sup> Disponible en: <https://data.jrc.ec.europa.eu/dataset/7342ea15-fd4f-4184-9603-98bd87d8239a>.

de salud pública; para servicios de policía; para el orden público y seguridad, y, finalmente, en el ámbito de la protección social. Sin embargo, tal y como puede observarse en el gráfico 1, los ámbitos donde actualmente existe una mayor presencia de inteligencia artificial son para las ayudas económicas extranjeras, para los servicios públicos, en asuntos económicos de otras industrias y para I+D en orden público y seguridad.

**Gráfico 7.** Número de algoritmos de inteligencia artificial que son utilizados en la Unión Europea, por funciones.



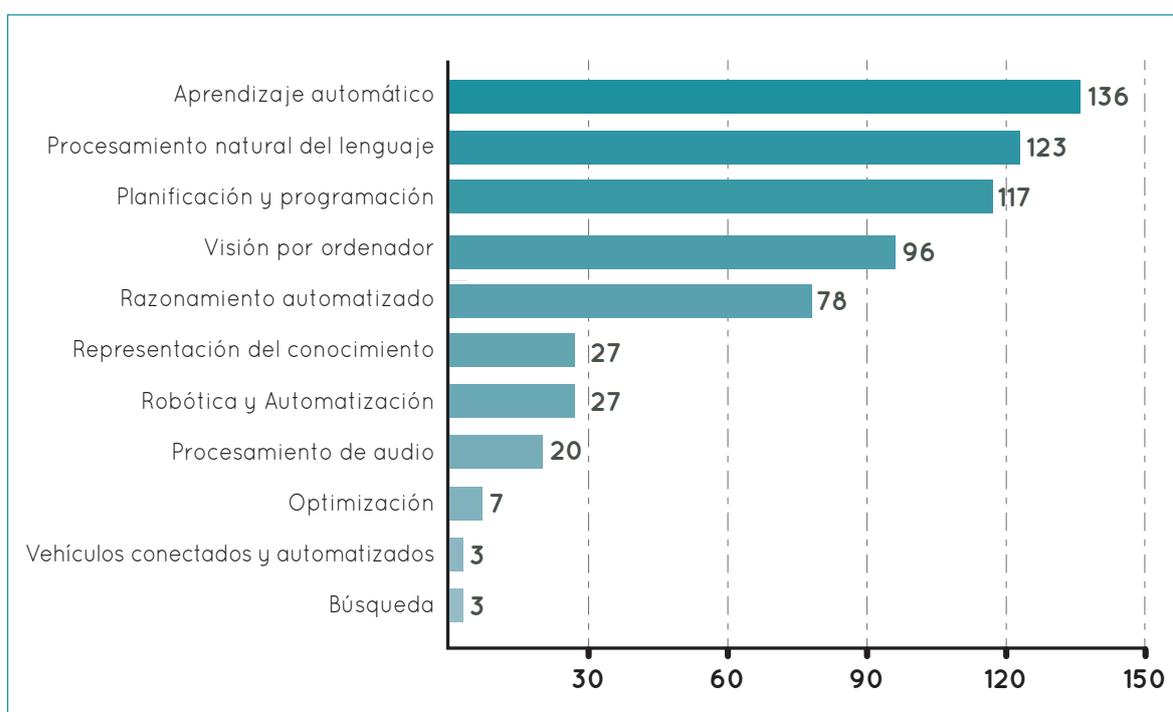


Fuente: elaboración propia a partir de los datos del *Joint Research Centre*.

En cuanto al tipo de IA más frecuente dentro del sector público se encuentra el *machine learning*, seguido del procesamiento de lenguaje natural y algoritmos para la planificación y programación (ver gráfico 8).

A su vez, estos algoritmos mayoritariamente son introducidos a nivel nacional (371 de un total de 686, que representan un 54,1 %), un 26,7 % son introducidos a nivel local y solo un 10,1 % y un 9,2 % son aplicados para funciones regionales o entre diferentes países, respectivamente.

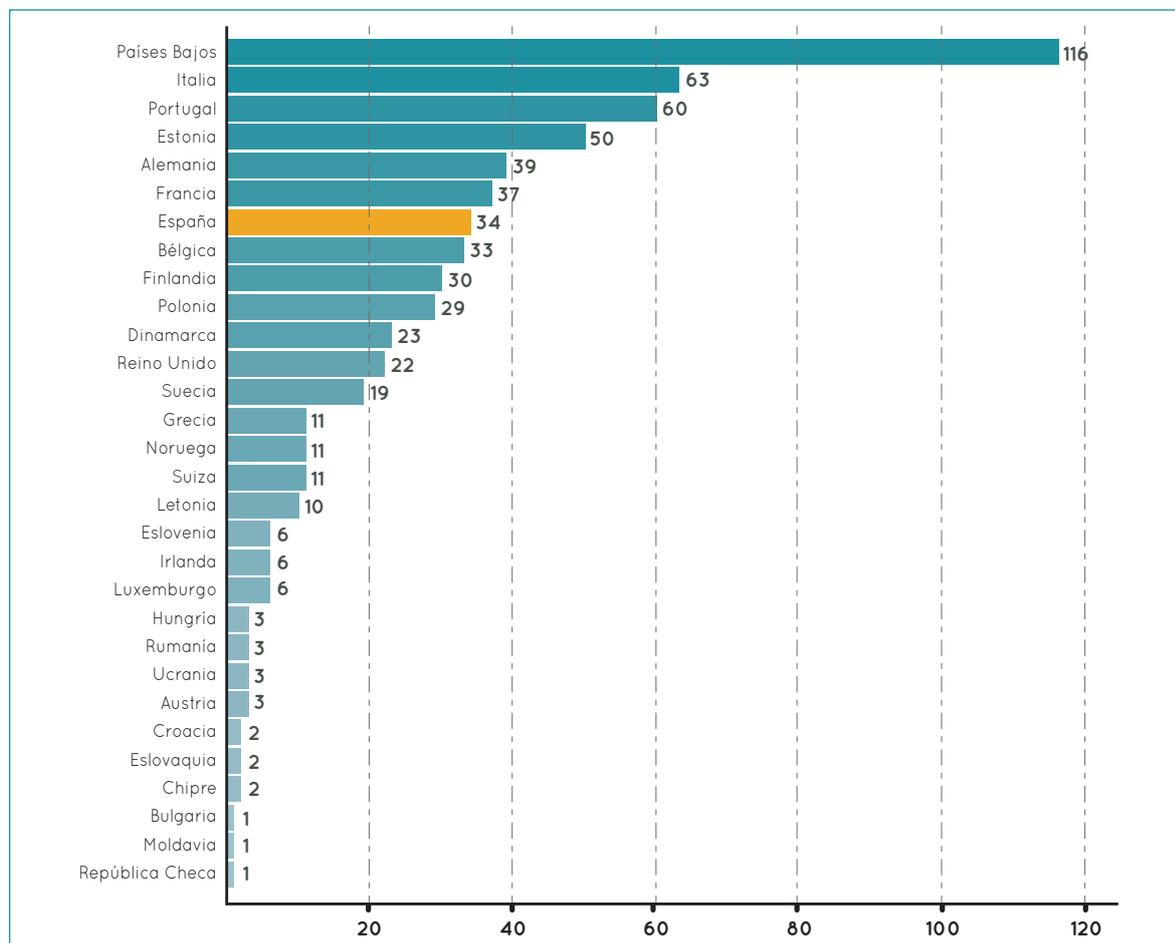
### Gráfico 8. Número de algoritmos de inteligencia artificial que son utilizados en el sector público dentro la Unión Europea, por tipo.



Fuente: elaboración propia a partir de los datos del *Joint Research Centre*.

Además, el siguiente gráfico muestra que los tres países donde hay más algoritmos introducidos, en desarrollo o en fase de prueba piloto en el sector público son Holanda (con 116 algoritmos), Italia (con 63 algoritmos) y Portugal (con 60 algoritmos). España se sitúa en el sexto lugar, muy por encima de la media.

**Gráfico 9.** Número de algoritmos de inteligencia artificial que son utilizados en la Unión Europea, por país.



Fuente: elaboración propia a partir de los datos del *Joint Research Centre*.

No obstante, cuando se consideran solamente los algoritmos actualmente implementados, el ranking varía. Así pues, teniendo en cuenta solo los 260 algoritmos implementados a lo largo del territorio de la Unión Europea, los países que actualmente trabajan con más algoritmos son Italia (con 38 algoritmos), Estonia (con 27) y Holanda (con 26).

De esta forma, la IA puede ser implementada en ámbitos específicos como salud, transporte público, defensa nacional, educación, administración de justicia o servicios sociales (ver sección 6.4). Así, la IA tiene un gran potencial para mejorar el trabajo de las administraciones públicas, capacidad que deberá demostrarse con los resultados de las aplicaciones llevadas a cabo en el presente y el futuro. Sin embargo, y tal y como se expondrá con mayor detalle en la sección 7, su introducción también está asociada a diferentes retos que el sector público debe intentar minimizar con el fin de poder maximizar sus beneficios. De ahí se desprende **la importancia de establecer un paso previo a la implementación de la IA que sirva para determinar si para aquel aspecto**

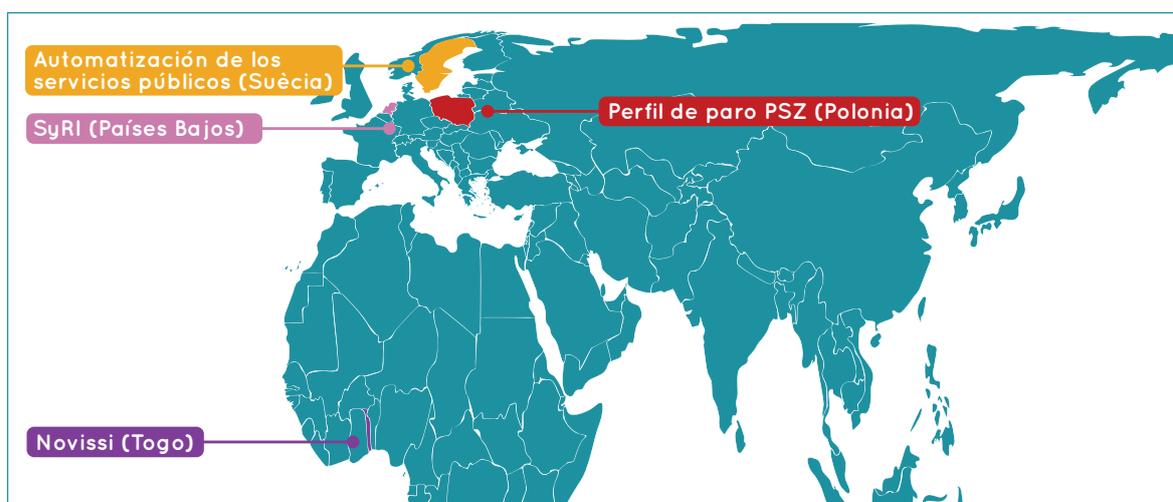
**concreto la IA es la mejor solución, mediante un análisis que valore tanto los potenciales beneficios como los riesgos que puede suponer.**

Para mostrar tanto los beneficios como los riesgos de la IA dentro de las administraciones públicas, a continuación se aportan casos de experiencias internacionales y nacionales que sirven de ejemplo de buenas y malas prácticas. A su vez, estos casos muestran la necesidad de realizar un esfuerzo adicional para garantizar que la IA sea utilizada de forma fiable, ética y equitativa, y, de esta forma, no impactar negativamente sobre la confianza de la ciudadanía en el Gobierno. De hecho, una investigación del Boston Consulting Group señala que el apoyo de la IA gubernamental está correlacionado con la confianza en el Gobierno, y que «la confianza en las instituciones es esencial para que los gobiernos tengan el apoyo necesario para desplegar las capacidades de la IA».<sup>11</sup> Esta confianza no depende únicamente de la eficiencia y la eficacia del servicio público, sino también de la satisfacción y de la justicia percibida en su funcionamiento.

### 6.3 Casos internacionales

Para ejemplificar cómo se están implementando sistemas de IA en procesos y servicios de la Administración pública, a continuación se expondrán cuatro casos diferentes que se han seleccionado por su popularidad y el volumen de información disponible. Además, se muestran dos casos con resultados muy positivos (Suecia y Togo) y otros dos donde han surgido problemáticas (Polonia y Países Bajos).

#### Gráfico 10.



Fuente: elaboración propia.

<sup>11</sup> Carrasco et al., 2019, citado en Berryhill, J., 2019

## Suecia (Automated Public Services)

 <b>Automatización de los servicios públicos en Trelleborg</b>	
<b>País</b>	Suecia
<b>Nombre</b>	<b>Automatización de los servicios públicos</b>
<b>Administración pública que lo ha implementado</b>	Protección social, a nivel local
<b>Impacto esperado</b>	Reducción del tiempo de espera, aumento de la eficiencia, mejora de la experiencia de la ciudadanía
<b>Descripción de la innovación</b>	Desde 2016, los servicios sociales del municipio de Trelleborg (Suecia) cuentan con un programa que utiliza IA para la automatización de diversas ayudas sociales. Ante un escenario donde los ciudadanos tenían que esperar una media de 8 días (que a veces llegaban a 20), y que acababa sobrecargando a los profesionales, que tenían que responder a sus consultas y gestionar las peticiones manualmente, la automatización robótica de procesos (RPA, por sus siglas en inglés) se planteó como la mejor solución para reducir tiempo de espera y retrasos en los pagos a los ciudadanos. Sin embargo, la supervisión de un profesional se ha mantenido para aquellos casos donde se rechaza la solicitud. En 2020, el sistema ya era capaz de procesar solicitudes de atención domiciliaria, prestaciones de enfermedad, prestaciones de paro e impuestos.
<b>Resultado e impacto</b>	<p>Este sistema automatizado ha permitido reducir el tiempo de espera en las solicitudes de ayudas significativamente, se ha reducido el tiempo de gestión de casos con vulnerabilidad económica y todas las decisiones de asistencia financiera se toman en un plazo de 24 horas.</p> <p>Además, un estudio sobre el impacto en los profesionales muestra un efecto positivo, dado que la inteligencia artificial los ayudaba a realizar su trabajo de forma más eficaz y les aportaba seguridad jurídica. Así pues, se constata el potencial de la colaboración tecnología y humano para redefinir las prácticas de asistencia social. No obstante, en otro informe se ha señalado la importancia de hacer que esa tecnología sea recibida con confianza por los profesionales, ya que, si no es así, se muestran reticentes a los cambios.</p> <p>Por lo tanto, este caso muestra cómo la automatización de procesos y los sistemas de inteligencia artificial pueden ser muy beneficiosos, siempre que no se descuide la relación de los profesionales con esos sistemas.</p>

## Togo (Novissi)

 <b>Novissi</b>	
<b>País</b>	Togo
<b>Nombre</b>	Novissi
<b>Administración pública que lo ha implementado</b>	Gobierno de Togo, nivel estatal
<b>Impacto esperado</b>	Acceso a medidas de emergencia del Gobierno, transferencias monetarias
<b>Descripción de la innovación</b>	<p>Ante la situación de urgencia marcada por la covid-19, el Gobierno de Togo tenía la voluntad de desplegar ayudas económicas para aquellas familias con más necesidades y, así, reducir el impacto negativo de la pandemia. No obstante, el Gobierno de Togo no disponía de un registro social completo que le permitiese identificar a las personas más pobres, y la pandemia hacía imposible la recogida de un nuevo registro. La falta de información aumentaba la exclusión de personas elegibles para las ayudas. Por ese motivo tuvieron que buscar una alternativa. Esa alternativa, llamada Novissi, se desarrolló e implementó mediante algoritmos de aprendizaje automático, datos satélites y redes de telefonía móvil. Inicialmente, las ayudas se distribuyeron a familias que cumpliesen con los siguientes tres requisitos: 1) Estar registrado en la plataforma de Novissi y haber facilitado información básica desde su teléfono móvil. 2) Estar registrado para votar en una región específica (Greater Lomé). 3) Autodeclarado trabajar en un empleo informal en su registro de votantes. Más adelante, se quiso extender este servicio a las personas en situación de pobreza en las zonas rurales del país.</p>
<b>Resultado e impacto</b>	<p>La evaluación de este programa muestra el potencial de utilizar nuevas fuentes de datos complementarias a las fuentes tradicionales para llegar a un mayor volumen de personas elegibles, sobre todo en contextos de crisis donde los datos tradicionales pueden faltar o no estar actualizados. De esta forma, se pone en relieve el potencial que tiene el machine learning para procesar grandes cantidades de datos y compaginar diversas fuentes de datos, que no sería posible de otro modo. En este caso concreto, por lo tanto, la evaluación muestra cómo la agilidad del sistema de <i>machine learning</i> permitió reducir errores y llegar a más personas que necesitaban la ayuda. Además, redujo el número de personas que habrían sido excluidas de forma incorrecta de la asistencia social en comparación con la opción geográfica.</p>

## Países Bajos (SyRI)

 SyRI	
País	Holanda
Nombre	SyRI (Systeem Risico Indicatie)
Administración pública que lo ha implementado	Protección social, a nivel central y municipal
Impacto esperado	Capacidades de inspección mejoradas, mejora del bienestar de los niños, reducción del mal uso de los fondos públicos
Descripción de la innovación	En 2012, la Agencia Tributaria de Holanda empezó a utilizar algoritmos de autoaprendizaje para crear perfiles de riesgo de cometer fraude con el fin de evitar el fraude en las ayudas destinadas al cuidado de niños.
Resultado e impacto	<p>Tras unos años de estar en funcionamiento, este sistema se retiró por tener consecuencias claramente negativas. El algoritmo se había desarrollado de tal forma que categorizaba como deudores a familias que habían rellenado incorrectamente los documentos de solicitud. Al mismo tiempo, tener una doble nacionalidad también influía en este perfilado, así como provenir de un nivel socioeconómico bajo, ser inmigrante o pertenecer a una minoría étnica, características que conducían al algoritmo a penalizar desproporcionadamente a estos grupos poblacionales. Además, los profesionales encargados de supervisar los casos dejaron pasar muchas falsas alarmas y siguieron con la solicitud de retorno del dinero.</p> <p>Como resultado, más de 10.000 personas entraron en situación de pobreza y otras se suicidaron tras recibir facturas de deudas con unas cuantías imposibles. Incluso más de 1.100 niños fueron separados de sus familias y llevados a centros de acogida. Un total de 30.000 familias se vieron afectadas por este algoritmo.</p> <p>Ante este escenario, se inició un proceso judicial para investigar el algoritmo, que condujo a una multa de 2,75 millones de euros a la Agencia Tributaria, y el Gobierno en funciones dimitió</p>

## Polonia (Unemployed Profiling)


**Perfil de paro PSZ (Publiczne Służby Zatrudnienia)**

<b>País</b>	Polonia
<b>Nombre</b>	PSZ (Publiczne Służby Zatrudnienia)
<b>Administración pública que lo ha implementado</b>	Asuntos económicos, nivel central y municipal
<b>Impacto esperado</b>	Servicios públicos personalizados, reducción del paro, mejora de la eficiencia
<b>Descripción de la innovación</b>	<p>En 2012, el Ministerio de Trabajo y Política Social de Polonia tenía la voluntad de reformar las oficinas laborales, que en aquel momento tenían poco personal y en las que las tareas desarrolladas eran poco eficientes. No obstante, el Gobierno de Polonia no disponía de suficientes recursos económicos para ampliar la plantilla. De esta forma, se recurrió a un sistema automatizado de perfilado para gestionar más eficientemente los casos de personas desempleadas. El sistema se diseñó con el objetivo de situar a las personas desempleadas en una de las tres categorías, teniendo en cuenta características individuales. Cada una de estas tres categorías está asociada al tipo de programa al cual la persona puede optar, es decir, define la elegibilidad de las personas ante el catálogo de servicios (como, por ejemplo, programas de inserción laboral, de formación profesional e incluso del subsidio de desempleo). Los datos –como la edad, el sexo y la duración del paro– se recogían durante una entrevista inicial, y después servían para categorizar a cada individuo. Por lo tanto, este sistema de asignación definía el nivel de apoyo y la carga de recursos a destinar a cada persona, donde una de las categorías era la ausencia de ayuda estatal.</p>
<b>Resultado e impacto</b>	<p>Después de 7 años, en 2019 se dictaminó la eliminación de este sistema de categorización, ya que se consideró que era ineficiente y, además, se identificaron patrones de discriminación: ser madre soltera, tener una discapacidad o vivir en una zona rural implicaban una peor categoría y, por lo tanto, recibir menos asistencia por parte de los servicios profesionales estatales. Además, un 44 % del personal que trabajaba en las oficinas afirmaba que no le era útil para reducir la carga de trabajo del día a día, y un 80 % consideraba que había que cambiar el sistema.</p>

## 6.4 Aplicaciones en el ámbito social

Tal y como se ha visto anteriormente, existen numerosos casos donde se ha empezado a experimentar e introducir la inteligencia artificial dentro de las administraciones públicas, para el desarrollo de tareas de carácter muy diverso. Sin embargo, cuando estas innovaciones se desarrollan en el ámbito social –y más concretamente en el de la protección social–, despiertan mayor interés y la necesidad de avanzar con prudencia, dado que pueden impactar de forma más notoria en los ciudadanos. Precisamente por su naturaleza, este ámbito es especialmente sensible, dado que las personas a las cuales va dirigida esa protección suelen encontrarse en situaciones de vulnerabilidad socioeconómica. De hecho, la protección social es «la cara más visible de los gobiernos», de forma que la introducción de la IA debe realizarse con más cuidado, ya que los riesgos pueden ser muy grandes si no se implementa bien.

Así pues, con una buena implementación de la IA (entendida tal y como está definida por la Unión Europea, en tanto que está centrada en las personas y es fiable), las administraciones públicas del ámbito social pueden utilizar esta herramienta para evaluar la elegibilidad y las necesidades, para la toma de decisiones de inscripción y para la asignación de prestaciones, así como para realizar el seguimiento de las mismas. A su vez, la IA puede permitir que la solicitud de una cita con un profesional sea con aquel que pueda responder mejor a las necesidades del ciudadano. También puede ayudar a distribuir mejor a los usuarios a lo largo del territorio, gestionar las listas de espera o conocer la evolución de la demanda en tiempo real.

Se recomienda leer el *informe Inteligencia artificial: decisiones automatizadas en Cataluña*, de la Autoridad Catalana de Protección de Datos. Pese a que expone ejemplos de aplicación de algoritmos en diversos campos, como la salud, el sistema judicial o la educación, también enumera hasta siete aplicaciones a nivel social. Aunque no se entrará a comentarlas todas, sí que se hará mención de ellas. El Ayuntamiento de Barcelona atiende a 50.000 personas de media al año a través de sus 40 centros de servicios sociales. Esas personas tienen problemáticas muy diversas, que son atendidas por una plantilla de 700 personas. El Ayuntamiento optó por explotar la información almacenada en las trescientas mil entrevistas que se habían realizado, y aplicando técnicas de machine learning, ha desarrollado un modelo capaz de analizar los nuevos casos, clasificar las demandas y aportar posibles respuestas.

Finalmente, para el seguimiento de las intervenciones, la IA permite que la atención para cada usuario sea más individualizada e integral (como, por ejemplo, con la historia social única electrónica del Sistema Cohesiona de la Junta de Andalucía).<sup>12</sup> También permite anticipar futuras demandas y, por lo tanto, actuar de forma preventiva. A continuación, se introducen dos casos de aplicaciones de algoritmos en el territorio español que han tenido resonancia a nivel estatal debido a la sensibilidad en el ámbito donde operan.

## España (VioGén)

	
<b>País</b>	<b>España</b>
<b>Nombre</b>	<b>VioGén</b>
<b>Administración pública que lo ha implementado</b>	Ministerio de Interior, a nivel estatal
<b>Impacto esperado</b>	Poder hacer seguimiento y proteger a las mujeres víctimas de violencia de género y a sus hijos/as en cualquier punto del territorio nacional
<b>Descripción de la innovación</b>	En 2007, el Ministerio de Interior del Gobierno de España desarrolló el Sistema de Seguimiento Integral de casos de Violencia de Género (VioGén). VioGén fue diseñado con el objetivo de determinar el nivel de peligro de sufrir nuevas agresiones, escoger los protocolos más adecuados para proteger a las víctimas y a sus hijos/as y establecer un seguimiento para determinar y ajustar las medidas de seguridad. No obstante, la decisión final es supervisada por un profesional, que puede subir el nivel de riesgo, pero nunca bajarlo.
<b>Resultado e impacto</b>	Desde su implementación hasta mayo de 2023, el sistema ha sido utilizado para valorar 6.047.700 casos, con 77.213 casos activos. Además, el sistema se ha modificado cinco veces para adaptarlo mejor a las nuevas situaciones. Sin embargo, este sistema no es inequívoco, y siguen habiendo casos de mujeres asesinadas por sus parejas o exparejas, incluso después de haber denunciado. A su vez, el algoritmo también es opaco y no se conoce cuál es el sistema de puntuación de riesgo que utiliza.

<sup>12</sup> Ver más información sobre el programa [aquí](#).

## España (BOSCO)

 <b>BOSCO</b>	
<b>País</b>	España
<b>Nombre</b>	BOSCO
<b>Administración pública que lo ha implementado</b>	Servicios sociales, a nivel estatal
<b>Impacto esperado</b>	Ofrecer una mejor regulación del otorgamiento de bonos sociales para el pago de la factura de la luz
<b>Descripción de la innovación</b>	El algoritmo BOSCO fue introducido por el Gobierno español e implementado por las eléctricas con la finalidad de regular el bono social sobre el pago de la factura de la luz. Este bono permite el acceso a un descuento en la factura de la luz.
<b>Resultado e impacto</b>	Pese a ser totalmente opaco para la población, hecho que ha sido denunciado por la Fundación Civio, este algoritmo también parece ser ineficiente, en tanto que personas elegibles que lo han solicitado finalmente no lo han obtenido. De este modo, pese a su opacidad, se ha manifestado la preocupación sobre la posibilidad de que este algoritmo esté discriminando a grupos poblacionales que tienen derecho a percibirlo, como serían jubilados y viudas.

## 7. BENEFICIOS Y RIESGOS DE LA APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL EN LAS ADMINISTRACIONES PÚBLICAS EN EL ÁMBITO SOCIAL

A lo largo de este informe, hemos ido viendo cuáles han sido los avances por parte de los gobiernos –con especial énfasis en España– para garantizar que las administraciones públicas y la sociedad en su conjunto puedan maximizar los beneficios y minimizar los riesgos que puede comportar un futuro con inteligencia artificial. Si bien se han ido ejemplificando algunos de ellos, aún no hemos profundizado sobre a qué beneficios y riesgos nos estamos refiriendo cuando hablamos de IA. Por lo tanto, en esta sección se presentan los beneficios y riesgos que han sido identificados hasta el momento.

Tal y como se ha mencionado, la introducción de la inteligencia artificial en las administraciones públicas comporta una serie de beneficios y riesgos asociados que hay que tener presentes, ya que pueden impactar directamente sobre el ejercicio de los derechos sociales.

### Beneficios

En cuanto a los **beneficios**, esta tecnología ofrece numerosas oportunidades al sector público para diseñar mejores políticas y mejorar la toma de decisiones, mejorar la comunicación y el compromiso con los ciudadanos y mejorar la velocidad y la calidad en la provisión de los servicios públicos, así como sustituir tareas rutinarias realizadas por los funcionarios públicos y permitir que estos realicen tareas de mayor valor añadido. De esta forma, la IA permite mejorar el diseño y la provisión de servicios públicos, así como mejorar la gestión interna de las instituciones estatales.

Respecto al primer punto, **mejorar el diseño y la provisión de servicios públicos**, la inteligencia artificial puede ayudar a las administraciones públicas a identificar intereses, preocupaciones y percepciones de diferentes actores para que sean incluidos en la agenda del Gobierno. Además, puede ayudar a detectar problemas en el acceso a determinados servicios, a comprender mejor el comportamiento de los ciudadanos y ver el impacto de diferentes entidades públicas sobre determinados grupos poblacionales.

A su vez, esta tecnología permite gestionar grandes cantidades de datos que generan los ciudadanos a través de su interacción con dispositivos

móviles y redes sociales. Mediante este tratamiento, los profesionales de la Administración pública pueden diseñar servicios más personalizados conforme a la realidad de la ciudadanía y, de ese modo, proveer una mejor atención a las entidades, con un mayor ahorro de tiempo y recursos.

Asimismo, la IA también ofrece la posibilidad de hacer que las interacciones con los ciudadanos sean más eficientes. A través de robots conversacionales, los ciudadanos pueden resolver dudas o solicitudes de forma ágil, lo que favorece una mejor satisfacción de los mismos, siempre que sepan que hay un algoritmo detrás.

Respecto al segundo punto, **mejorar la gestión interna de las instituciones estatales**, la inteligencia artificial dispone de técnicas de predicción, optimización o control que pueden utilizarse para ayudar a la asignación y gestión de recursos económicos y la detección y prevención de fraude y para evitar usos ineficientes de estos recursos. A su vez, puede ayudar a la generación de nueva regulación y su actualización y a garantizar su cumplimiento. Gracias a su capacidad de automatización, esta tecnología también puede utilizarse para tareas repetitivas y rutinarias, permitiendo a los profesionales de las administraciones públicas realizar actividades más complejas que requieren mayores habilidades emocionales, de creatividad y de perspectiva humana. Así pues, uno de los ejemplos donde puede automatizarse el servicio público es en la entrada de datos a través del reconocimiento automático de la escritura manual, el reconocimiento de voz o el procesamiento del lenguaje natural.

De esta forma, al reducir la carga de trabajo eliminando las tareas repetitivas y mecánicas, la Administración gana tiempo, que podrá dedicar a prestar una atención más rápida, eficiente y personalizada y a mejorar otros procesos que lo requieran o proveer nuevos servicios.

Finalmente, las capacidades de precisión y predicción pueden ser utilizadas por las administraciones públicas para realizar diagnósticos o definir líneas de actuación ante una emergencia.

## Riesgos

En cuanto a los **riesgos** de aplicar la IA en el sector público, se encuentra en primer lugar la **falta de accountability** (rendición de cuentas en castellano), **transparencia y explicabilidad de los algoritmos**.

La OCDE define los principios de *accountability*, transparencia y explicabilidad de la siguiente forma:

- ▶ **Principio de *accountability*:** los actores de la IA deben garantizar que sus sistemas de IA son fiables.
- ▶ **Principio de transparencia y explicabilidad:** los actores deben comprometerse con la transparencia y la divulgación responsable de los sistemas de IA. Para ello, deben proporcionar información significativa adecuada al contexto y coherente con el estado de la cuestión.

Las principales críticas de esta tecnología giran en torno a su **opacidad para ser entendida por los ciudadanos**, donde la complejidad de su proceso juega un papel relevante. Diversos casos internacionales han mostrado cómo, a menudo, estos algoritmos de inteligencia artificial no son de código abierto (y, por lo tanto, no son transparentes). Consiguientemente, la falta de transparencia dificulta que puedan auditarse los algoritmos y, por lo tanto, que pueda cumplirse con el principio de *accountability*.

A su vez, diversos algoritmos tampoco son fácilmente comprensibles por parte de los ciudadanos, incumpliendo el principio de transparencia y explicabilidad. De acuerdo con este segundo principio, **las personas que utilizan estas soluciones han de poder entender cómo se ha llegado a ese resultado, y tanto estas como las personas que no las utilizan pero se ven afectadas han de poder ejercer su derecho a reclamar si no están conformes**. Sin embargo, la opacidad en el código, como de las cajas negras, impide que pueda cumplirse este principio. Además, este aspecto está estrechamente vinculado con el principio de *accountability*, ya que sin una garantía de transparencia y explicabilidad, demostrar que un sistema IA es fiable resulta un ejercicio mucho más complejo.

Estos principios, pues, son esenciales para garantizar que los algoritmos no van en contra del bien común y que, por lo tanto, no se están vulnerando los derechos sociales de la ciudadanía. Precisamente, entender cómo se ha obtenido un resultado es esencial para identificar sesgos o patrones de discriminación. De esta forma, cuando los actores públicos no garantizan un cumplimiento de estos principios, el control público queda reducido, lo que afecta especialmente a los ciudadanos pertenecientes a minorías y a grupos desfavorecidos, en tanto que no pueden exigir la garantía de sus derechos sociales. Con el fin de evitarlo, están surgiendo iniciativas basadas en la creación de registros de algoritmos en ciudades como Bar-

celona,<sup>13</sup> con el objetivo de dotar estos sistemas de mayor transparencia y fiabilidad.

Aparte de la falta de cumplimiento de estos principios, otro problema asociado a la **amplificación de sesgos y discriminación** es la falta de datos de buena calidad. A menudo, la generación de estos patrones de discriminación surge de un aprendizaje que realizan los algoritmos a partir de los datos con los que se entrenan y aprenden. De esta forma, si los datos introducidos no son de calidad y contienen errores y discriminaciones ya existentes dentro de la sociedad, estos algoritmos simplemente serán una réplica de esos sesgos. Por lo tanto, de ahí se deriva la gran importancia de disponer de datos de calidad y que no conduzcan a la perpetuación de desigualdades del *Matthew effect*, donde los privilegiados obtienen ventajas, mientras que los que ya se encuentran en una situación de desventaja aún sufren más.<sup>14</sup> A su vez, estos sesgos pueden fomentar un proceso de polarización social y adoptar comportamientos radicales sobre determinados grupos poblacionales.

Otro riesgo a tener en cuenta es la **seguridad y privacidad de la información utilizada**. La información utilizada por las administraciones públicas es extremadamente sensible, y la fuga de información o los ataques cibernéticos podrían comportar graves consecuencias, tanto para las entidades públicas como para los propios ciudadanos. En este sentido, también cabe señalar que la Ley de Protección de Datos<sup>15</sup> puede suponer un impedimento para el entrenamiento de ciertos modelos si no se garantiza la seguridad y privacidad de la información utilizada.

En consecuencia, todos estos riesgos pueden acabar afectando la efectividad de las políticas públicas, dado que las posibles situaciones de exclusión de población diana podrían reducir el impacto de las mismas. Al fin y al cabo, todos estos riesgos impactan negativamente sobre la confianza depositada en las administraciones públicas por parte de los ciudadanos. El conjunto de errores y perpetuación de desigualdades acaba desgastando la confianza de la población sobre la eficacia de la actuación de las administraciones públicas, con implicaciones directas sobre la voluntad de participación ciudadana y, finalmente, sobre el estado de la democracia.

En resumen, entre los principales beneficios de la IA se encuentran **una mejora en la eficiencia de los recursos** (tanto económicos como de personal), **una mejor provisión de un servicio público y una mejor adecuación**

---

<sup>13</sup> Ver más información [aquí](#).

<sup>14</sup> Herzog, 2021, citado en OCDE, 2023, p.8

<sup>15</sup> Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

de las políticas públicas y programas de la agenda. En cambio, los riesgos se vinculan principalmente con la actual **falta de transparencia de estos procesos, la falta de datos adecuados para evitar situaciones de sesgos y discriminación, la sensibilidad de los datos tratados para garantizar su privacidad**, las **situaciones de falsos positivos y errores que conducen a situaciones graves para las personas afectadas** y, en consecuencia, finalmente, **la pérdida de la confianza de los ciudadanos sobre el trabajo realizado por el sector público**.

Gráfico 11.



## 8. CONCLUSIÓN

La inteligencia artificial tiene potencial para transformar significativamente nuestra sociedad. Cada vez más empresas y administraciones públicas integrarán en sus productos y procesos sistemas que utilicen inteligencia artificial, buscando que sean más eficientes, efectivos o beneficiosos.

La Unión Europea y los gobiernos nacionales llevan años trabajando en una regulación que constriña las posibilidades de desarrollo con la esperanza de minimizar y prevenir los riesgos sin afectar a la innovación y el avance en el campo de la IA. Una regulación que, *a priori*, será definitiva a finales de 2023.

Ante este contexto, de presente y futuro, es fundamental que las entidades del tercer sector fijen una posición y discurso común y lo trasladen tanto a las esferas políticas como públicas. En los próximos años, probablemente se promoverán la mayor parte de las regulaciones, pero también se consolidará la cosmovisión sobre la inteligencia artificial. Por lo tanto, **es fundamental que el tercer sector sea un actor con voz (y voto), si quiere participar en la conformación de la sociedad futura.**

Las entidades del tercer sector actualmente se encuentran en un momento clave para adoptar una estrategia proactiva y hacer escuchar su voz durante la elaboración de las próximas legislaciones. Asimismo, las entidades han de poder ejercer de puente entre la ciudadanía y las administraciones públicas para concienciarlas de los avances que tienen a su disposición e informarlas sobre cómo pueden beneficiarse de estos, mientras evitan todos sus riesgos.

Adicionalmente, contar con una visión común como tercer sector fortalece la acción desarrollada en el control de estas aplicaciones en el ámbito de los derechos sociales y, a la vez, permite impulsar nuevos acuerdos con diferentes agentes para poder aprovechar las oportunidades de esta tecnología.

En definitiva, **el tercer sector se encuentra ante una oportunidad única para, a través de una posición común, fortalecer su voz y seguir trabajando en la garantía de los derechos sociales en este nuevo entorno cambiante.**

## REFERENCIAS

- Ahn, M. J., & Chen, Y.-C. (2020). Artificial Intelligence in Government: Potentials, Challenges, and the Future. In S.-J. Eom & J. Lee (Eds.), *Dg.o '20: The 21st Annual International Conference on Digital Government Research* (pp. 243-252). Association for Computing Machinery. <https://doi.org/10.1145/3396956.3398260>
- AI Watch (2022). Spain AI Strategy Report. Comisión Europea. Consultado el día 21/06/2023. Disponible en: [https://ai-watch.ec.europa.eu/countries/spain/spain-ai-strategy-report\\_en](https://ai-watch.ec.europa.eu/countries/spain/spain-ai-strategy-report_en)
- ADB (2020). AI in Social Protection - Exploring Opportunities and Mitigating Risks. Recuperado de: <https://www.adb.org/publications/ai-social-protection-exploring-opportunities-mitigating-risks>
- Abu Elyounes, D. (10 de febrer de 2021). Why the resignation of the dutch government is a good reminder of how important it is to monitor and regulate algorithms. Medium. Recuperado el 25 de julio de 2023, de <https://medium.com/berkman-klein-center/why-the-resignation-of-the-dutch-government-is-a-good-reminder-of-how-important-it-is-to-monitor-2c599c1e0100>
- Algoritm Watch; Bertelsmann Stiftung. (Gener de 2019). Bertelsmann Stiftung. Automating Society. Taking Stock of Automated Decision-Making in the EU. Berlín: AW AlgorithmWatch gGmbH, gener de 2019. Recuperado de: [https://algorithmwatch.org/de/wp-content/uploads/2019/02/Automating\\_Society\\_Report\\_2019.pdf](https://algorithmwatch.org/de/wp-content/uploads/2019/02/Automating_Society_Report_2019.pdf)
- Belinchón, F. (2023). La primera ley europea sobre inteligencia artificial se amplía para incluir a ChatGPT y entra en su recta final. El País. Disponible en: <https://cincodias.elpais.com/economia/2023-06-14/la-primer-ley-europea-sobre-inteligencia-artificial-se-amplia-para-incluir-a-chatgpt-y-entra-en-su-recta-final.html>
- Berryhill, J. *et al.* (2019), 'Hello, World: Artificial intelligence and its use in the public sector', OECD Working Papers on Public Governance, No. 36, OECD Publishing, Paris, <https://dx.doi.org/10.1787/726fd39d-en>.
- Big Data Magazine (2022). Algoritmos de la Administración Pública española: poco conocidos, opacos y sin auditar. Recuperado el 25 de julio de 2023, de: <https://bigdatamagazine.es/algoritmos-de-la-administracion-publica-espanola-poco-conocidos-opacos-y-sin-auditar>

- Blázquez, L. (20 de febrer de 2023). ¿Qué es VioGén y cómo funciona?. Rtve. Recuperado el 25 de julio, de: <https://www.rtve.es/noticias/20230220/objetivo-igualdad-funcionamiento-viogen/2424326.shtml>
- Boucher, P. N. (2020). Artificial intelligence: How does it work, why does it matter, and what can we do about it?. Scientific Foresight Unit (STOA), European Parliament.
- Carrasco, M., S. Mills, A. Whybrew y A. Jura (2019), 'The citizen's perspective on the use of AI in government', BCG, 1 de marzo, [www.bcg.com/publications/2019/citizen-perspective-use-artificial-intelligence-government-digital-benchmarking.aspx](http://www.bcg.com/publications/2019/citizen-perspective-use-artificial-intelligence-government-digital-benchmarking.aspx)
- De Schutter, O. (2022). La no percepción de derechos en el contexto de la protección social. Consejo de Derechos Humanos. Obtenido de: <https://documents-ddsny.un.org/doc/UNDOC/GEN/G22/322/20/PDF/G2232220.pdf>
- European Commission (2019). Building Trust in Human-Centric Artificial Intelligence. Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions. COM(2019) 168. Brussels. Disponible en: <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence>
- Felton, J. (11 de maig de 2022). The ai that led to children being re-homed and the fall of an elected government. IFL Science. Recuperado el 25 de julio de 2023, de <https://www.iflscience.com/the-ai-that-led-to-children-being-rehomed-and-the-fall-of-an-elected-government-63622>
- Gleb Papyshv & Masaru Yarime (2023) The state's role in governing artificial intelligence: development, control, and promotion through national strategies, Policy Design and Practice, 6:1, 79-102, DOI: 10.1080/25741292.2022.2162252
- Heikkilä, M. (29 de marzo de 2022). Dutch scandal serves as a warning for europe over risks of using algorithms. Politico. Recuperado el 25 de julio de 2023, de <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/>
- High-Level Expert Group on Artificial Intelligence (2019). Ethics Guidelines for Trustworthy AI. European Commission. Disponible en: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

- Jiménez Arandia, P. (2023). Govern Obert: Transparencia algorítmica en el sector públic. (rep.) Vol.9. Generalitat de Catalunya. Generalitat de Catalunya. Recuperado de: [https://governobert.gencat.cat/web/content/01\\_Que\\_es/04\\_Publicacions/colleccio\\_govern\\_obert/GovernObert\\_9/Govern-Obert-9-cast.pdf](https://governobert.gencat.cat/web/content/01_Que_es/04_Publicacions/colleccio_govern_obert/GovernObert_9/Govern-Obert-9-cast.pdf)
- KOSA AI (2021). The importance of AI governance and 5 key principles for its guidance. Medium. Disponible en: <https://kosa-ai.medium.com/the-importance-of-ai-governance-and-5-key-principles-for-its-guidance-219798c8f407>
- Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications policy*, 44(6), 101976.
- La Moncloa (19 de maig de 2023). El Sistema VioGén supera los seis millones de valoraciones de riesgo a víctimas de violencia de género. Recuperado el 25 de julio de 2023, de: <https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/interior/Paginas/2023/190523-sistema-viogen-victimas-violencia-genero.aspx>
- Lokshin, M., & Umapathi, N. (23 de febrero de 2022). AI for social protection: mind the people. Brookings Institution. Recuperado el 28 de junio de 2023, de <https://www.brookings.edu/articles/ai-for-social-protection-mind-the-people/>
- Marilín Gonzalo (2023). Ley de IA en la Unión Europea: estas son las claves de la propuesta. Newtral. Disponible en: <https://www.newtral.es/ley-ia-union-europea-inteligencia-artificial-europa/20230520/>
- Minguíjon J. y Serrano-Martínez C. (2022). La Inteligencia Artificial en los Servicios Sociales: estado de la cuestión y posibles desarrollos futuros. *Cuadernos de Trabajo Social*, 35(2), 319-329. <https://doi.org/10.5209/cuts.78747>
- Misuraca, G., & Van Noordt, C. (2020). AI Watch-Artificial Intelligence in public services: Overview of the use and impact of AI in public services in the EU. JRC Research Reports, (JRC120399).
- Misuraca, G., i van Noordt, C. (2020). Overview of the use and impact of AI in public services in the EU, EUR 30255 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-19540-5, doi:10.2760/039619, JRC120399
- Ministerio de Asuntos Económicos y Transformación Digital (2020). Estrategia Nacional de Inteligencia Artificial. Disponible en: <https://portal.mineco.gob.es/es-es/digitalizacionIA/Paginas/ENIA.aspx>

- Niklas, J. (16 de abril de 2019). Poland: Government to scrap controversial unemployment scoring System. Algorithm Watch. <https://algorithmwatch.org/en/poland-government-to-scrap-controversial-unemployment-scoring-system/>
- NL Times. (17 de diciembre de 2020). Parents faced ‘unprecedented injustice’ for years in childcare subsidy scandal. NL Times. Recuperado el 25 de julio de 2023, de <https://nltimes.nl/2020/12/17/parents-faced-unprecedented-injustice-years-childcare-subsidy-scandal>
- OASI (s/d). Discriminación algorítmica. Recuperado de: <https://eticas-foundation.org/oasi/social-impact-algorithms/>
- Observatorio Nacional de Tecnología y Sociedad (2023). Uso de inteligencia artificial y big data en las empresas españolas. Secretaría de Estado de Digitalización e Inteligencia Artificial, Ministerio de Asuntos Económicos y Transformación Digital. Disponible en: <https://www.ontsi.es/es/publicaciones/uso-de-inteligencia-artificial-y-big-data-en-las-empresas-espanolas>
- OECD (2023). Global Trends in Government Innovation 2023, OECD Public Governance Reviews, OECD Publishing, Paris, <https://doi.org/10.1787/0655b570-en>.
- OECD (2021). ‘Tools for trustworthy AI: A framework to compare implementation tools for trustworthy AI systems’, OECD Digital Economy Papers, No. 312, OECD Publishing, Paris, <https://doi.org/10.1787/008232ec-en>.
- OECD (s/d). Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449
- ONTSI (2022). Uso de inteligencia artificial y big data en las empresas españolas. Disponible en: <https://www.ontsi.es/es/publicaciones/uso-de-inteligencia-artificial-y-big-data-en-las-empresas-espanolas>
- Our World in Data (2022). Artificial intelligence is transforming our world – it is on all of us to make sure that it goes well. Disponible en: <https://ourworldindata.org/ai-impact>
- Our World in Data (2022). The brief history of artificial intelligence: The world has changed fast – what might be next?. Disponible en: <https://ourworldindata.org/brief-history-of-ai>
- Ranerup, A., & Henriksen, H. Z. (2022). Digital Discretion: Unpacking Human and Technological Agency in Automated Decision Making in Sweden’s Social Services. Social Science Computer Review, 40(2), 445–461. <https://doi.org/10.1177/0894439320980434>

- Roser, M. (2022a). Artificial intelligence is transforming our world — it is on all of us to make sure that it goes well. Our World in Data. Disponible en: <https://ourworldindata.org/ai-impact>
- Roser, M. (2022b). The brief history of artificial intelligence: The world has changed fast - what might be next?. Our World in Data. Disponible en: <https://ourworldindata.org/brief-history-of-ai>
- Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36(2), 368–383. <https://doi.org/10.1016/j.giq.2018.09.008>
- Tangi L., van Noordt C., Combetto M., Gattwinkel D., Pignatelli F. (2022). AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector, EUR 31088 EN, Publications Office of the European Union, Luxembourg , ISBN 978-92- 76-53058-9, doi:10.2760/39336, JRC129301.
- Tangi, L., van Noordt, C., Combetto, M., Gattwinkel, D., & Pignatelli, F. (2022). AI Watch. European landscape on the use of Artificial Intelligence by the Public Sector (No. JRC129301). Joint Research Centre.
- Turkish Policy Quarterly. (7 de marzo de 2022). AI and democracy: Towards better public oversight. *Turkish Policy Quarterly*. <http://turkishpolicy.com/article/1108/ai-and-democracy-towards-better-public-oversight>
- Vélez, M. I., Gómez Santamaría, C., & Osorio Sanabria, M. A. (2022). Conceptos fundamentales y uso responsable de la inteligencia artificial en el sector público. Informe 2. Caracas: CAF. Recuperado de: <https://scioteca.caf.com/handle/123456789/1921>
- Veale, M., Matus, K., & Gorwa, R. (2023). AI and Global Governance: Modalities, Rationales, Tensions. *Annual Review of Law and Social Science*, 19.
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial Intelligence and the Public Sector—Applications and Challenges. *International Journal of Public Administration*, 42(7), 596–615. <https://doi.org/10.1080/01900692.2018.1498103>

DOCUMENTO 2

**RADAR DE ALGORITMOS  
Y PROCESOS  
DE DECISIÓN  
AUTOMATIZADA PARA  
EL ACCESO A LOS  
DERECHOS SOCIALES  
DE LA CIUDADANÍA**

DICIEMBRE DE 2023

# 1. INTRODUCCIÓN

El primer documento de este proyecto se centraba en la elaboración del informe *La inteligencia artificial en las administraciones públicas*. Este informe surgió con el objetivo de producir un documento que ayudase a las entidades del tercer sector a comprender mejor el fenómeno de la inteligencia artificial (IA), tanto en su vertiente histórica, técnica y de gobernanza como de aplicación actual, especialmente en el ámbito público y social. Por lo tanto, se trató de un informe divulgativo dirigido especialmente al tercer sector o a un público no especializado.

En este segundo documento del proyecto, el objetivo prioritario ha sido la identificación de sistemas de inteligencia artificial y procesos automatizados en el ámbito social que estuviesen siendo utilizados, o en fase piloto, por parte de las administraciones públicas catalanas.

La IA está en el centro del debate actual. Los últimos lanzamientos, y los que están por llegar, dejan entrever una potencial transformación de las relaciones sociales y el mercado laboral de alcance mundial. Sin duda alguna, **desde 2022 estamos viviendo en un avance de publicaciones de modelos y herramientas sorprendentes, sin demasiado tiempo para adaptarnos ni reflexionar sobre sus consecuencias.**

**El objetivo de este proyecto es la realización de una recopilación –también llamada radar o repositorio– de todos aquellos algoritmos, modelos de inteligencia artificial y/o procesos de decisiones automatizadas que se estén utilizando en Cataluña en el ámbito del acceso a los derechos sociales.**

**El radar se dirigirá a ciudadanos, administraciones públicas y entidades del tercer sector** con la finalidad de proporcionar información clara y completa sobre el impacto de esta tecnología y el acceso a los derechos sociales en Cataluña. Esto es muy relevante, dado que **el uso** que se pretende hacer **del repositorio determinará decisiones como qué información recoger o qué lenguaje utilizar, entre otras.**

La creación de este **radar** se realizó en tres pasos. En primer lugar, se diseñó el radar en sí mismo. Es decir, qué variables incluiría y, por lo tanto, qué información se recogería para cada algoritmo. Este ejercicio se realizó tomando como referencia otros repositorios internacionales.

En segundo lugar, se llevó a cabo el trabajo de campo. Después de diseñar un cuestionario para una entrevista e implementar un formulario en línea, se contactó directamente con responsables de las administraciones públi-

cas que se consideraba que podrían contar con información. A estas personas se les dio la opción de rellenar directamente el formulario o realizar una entrevista. Para ampliar las posibilidades de respuestas, finalmente se decidió hacer público el formulario y difundirlo a través de las redes sociales.

En tercer lugar, después de la recogida de respuestas, se procesó, limpió y completó la información para configurar el radar de algoritmos.

La principal dificultad de la creación de este radar es el desconocimiento del universo poblacional. En otras palabras, **se desconoce el número total de sistemas de IA actualmente implementados o en proceso de desarrollo que están siendo utilizados por las administraciones públicas catalanas.** No obstante, en las entrevistas con responsables de diversas administraciones se expresó con claridad el hecho de que, en el ámbito de los derechos sociales, existen pocos casos de uso de estos algoritmos. **En total se han mapeado hasta 12 algoritmos de 4 administraciones diferentes.**

No obstante, **el radar de algoritmos es la primera fase de un proyecto más amplio.** En esta fase, el objetivo primordial ha sido mapear todos los algoritmos que estaban siendo utilizados.

Así pues, este documento se estructura en tres apartados:

- (1)** Explicar la metodología seguida para recabar toda la información sobre los algoritmos utilizados por la Administración pública de Cataluña.
- (2)** Explicar las variables que integran el radar, es decir, los campos en el ámbito social para los que se recabó información.
- (3)** Presentar las fichas elaboradas para cada uno de los 12 algoritmos recogidos.

## 2. METODOLOGÍA

Esta fase ha consistido en el mapeo de los diferentes algoritmos de IA y procesos de decisión automatizada que son utilizados por las administraciones públicas catalanas. Dado que son sistemas bastante recientes y poco conocidos, para realizar este mapeo se ha tenido que recurrir a una metodología combinada. Concretamente, el equipo investigador ha creado un formulario donde se incluyen todos los campos y variables de interés, a la vez que se han realizado entrevistas a personas identificadas como relevantes.

Así pues, en primer lugar, se han identificado un total de 28 personas. Estas personas se consideran relevantes por cumplir dos requisitos: en primer lugar, trabajar en administraciones públicas catalanas y, en segundo lugar, poseer conocimiento directo o haber participado en el diseño o implementación de un algoritmo que estuvo o está en uso en la Administración pública. Por lo tanto, estas personas son la fuente de información principal del radar.

Estas personas fueron identificadas en gran medida gracias al conocimiento de campo de la Taula d'entitats del Tercer Sector, junto con la cual se elaboró un listado de personas que potencialmente podrían tener conocimiento de casos de uso. Algunas de estas personas aportaron contactos sobre otros responsables que previamente no habían sido identificados. Bajo esta dinámica de bola de nieve, se acabó por configurar el listado definitivo de personas contactadas. Tal y como puede observarse en el siguiente gráfico, la mayoría de estos perfiles trabajan en el Ayuntamiento de Barcelona (un 28 %), en la Generalitat de Cataluña (un 21 %) y en diferentes fundaciones (un 21 %).

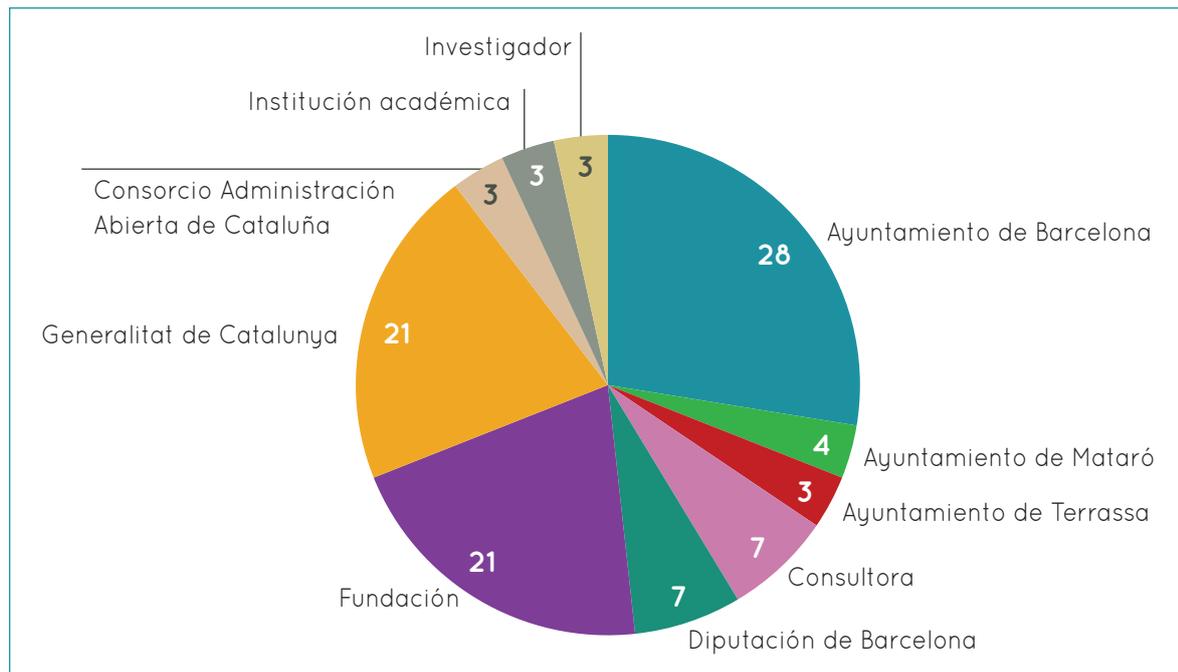
Una vez identificados estos perfiles, se les ha contactado para obtener su colaboración. En este contacto, se han ofrecido dos opciones para facilitar la información al equipo investigador: rellenando el [formulario](#) y/o concertando una entrevista para resolver dudas o profundizar en cuestiones concretas vinculadas a la información a recoger.

Paralelamente a la difusión a este grupo de perfiles identificados como relevantes, se ha ampliado la difusión del formulario por las redes sociales con el objetivo de maximizar el número de algoritmos recogidos y, por lo tanto, elaborar un repositorio lo más completo posible.

En total se ha obtenido un total de 13 respuestas al formulario, y se han realizado diversas entrevistas con diferentes personas e instituciones. En

concreto, se han organizado entrevistas con profesionales que trabajan en el Ayuntamiento de Barcelona, el Instituto Municipal de Servicios Sociales de Barcelona, la Generalitat de Catalunya, la Administración Abierta de Catalunya (AOC) y la Fundación TIC Salud Social.

### Gráfico 12. Organizaciones donde trabajan las personas contactadas



La información recopilada en el radar ha sido proporcionada de forma mayoritaria por las personas contactadas que rellenaron el formulario. No obstante, el sistema DPR fue registrado por tres expertos distintos, por lo que sus respuestas fueron combinadas en una sola. El registro de RisCanvi era muy breve, por lo que se rehízo por completo a partir de la información pública encontrada. Para el resto de sistemas se hicieron modificaciones o ajustes menores en la redacción.

## 3. EL DISEÑO DEL RADAR DE ALGORITMOS

El radar o repositorio se ha diseñado teniendo en cuenta tanto referencias internacionales como decisiones del equipo investigador de KSNET. Entre las primeras, poseen una gran relevancia el repositorio Casos de IA seleccionadas del sector público de AI-WATCH, del Joint Research Centre de la Comisión Europea,<sup>16</sup> y OASI, de la Fundación Ética.<sup>17</sup> Aunque en menor medida, también se han tenido en cuenta los registros de IA de Ámsterdam y Helsinki.<sup>18</sup>

Los dos primeros casos mencionados consisten en bases de datos para registrar la información recogida de los diversos algoritmos, donde el primero de estos –AI-WATCH, del Joint Research Centre de la Comisión Europea– se centra en el territorio europeo y el segundo –OASI, de la Fundación Ética– incluye casos internacionales más allá del territorio europeo.

De esta forma, para la selección de las variables del radar de algoritmos se ha buscado la comparabilidad con estas dos bases de datos, incorporando campos comunes como son los de objetivo, estatus y beneficios (Casos de IA) y riesgos y auditoría (OASI).

En la siguiente tabla se recogen los campos incluidos en el radar, que fueron rellenados por los expertos contactados a través del formulario en línea o por KSNET a partir de la información aportada por los expertos durante la entrevista.

### 3.1 Campos del radar de algoritmos

A continuación, se detallan los campos del radar. La información para cada campo se recogió mediante una pregunta específica del cuestionario.

#### Bloque 1. Información básica

##### 1. Nombre

En caso de tenerlo, nombre interno o público del algoritmo.

##### 2. Descripción

Contexto en el cual se desarrolla y se aplica este proceso automatizado/ algoritmo/sistema de IA. Es decir, qué problemática se había detectado,

<sup>16</sup> Enlace: <https://data.jrc.ec.europa.eu/dataset/7342ea15-fd4f-4184-9603-98bd87d8239a>.

<sup>17</sup> Enlace: <https://eticas.tech/oasi>.

<sup>18</sup> Amsterdam: <https://algoritmeregister.amsterdam.nl/>. Helsinki: <https://ai.hel.fi/en/ai-register/>.

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.

### 3. Objetivo

¿Cuál es el objetivo del proceso automatizado/algoritmo/sistema de IA?  
¿Para qué función ha sido desarrollado?

#### 3.1 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.

<b>Automatización de tareas</b>	Realizar de forma automatizada un conjunto de tareas que a una persona llevaría mucho más tiempo realizar.
<b>Recopilación de datos personales</b>	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.
<b>Evaluación del comportamiento humano</b>	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.
<b>Reconocimiento de rasgos faciales</b>	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.
<b>Identificación de imágenes de caras</b>	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.
<b>Predicción del comportamiento humano</b>	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.
<b>Elaboración de perfiles y clasificación de personas</b>	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.
<b>Simulación del habla humana</b>	Generar un habla que se parezca mucho a la forma como hablan las personas para propósitos públicamente conocidos o desconocidos.
<b>Reconocimiento de imágenes</b>	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 datos públicamente conocidos o desconocidos.

<b>Generación de traducciones automáticas</b>	Traducir automáticamente textos o discursos escritos de un idioma a otro/s.
<b>Generación de resultados de búsqueda en línea</b>	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.
<b>Reconocimiento de sonidos</b>	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.
<b>Xatbot conversacional</b>	Proporcionar de forma automatizada respuestas a las consultas ciudadanas.

## Bloque 2. Desarrollo e implementación

### 4. Administración responsable

Administración pública que está utilizando o utilizará el algoritmo.

### 5. 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.

### 6. Estatus

¿En qué situación se encuentra la implementación?

- ▶ 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.
- ▶ Otro: ...

### 7. 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.

## 7.1 Fecha de finalización

En caso de que el algoritmo esté en desuso, ¿en qué año dejó de utilizarse? Si está bajo prueba piloto, ¿cuándo finaliza la prueba?

## Bloque 3. Colectivo, beneficios y riesgos

### 8. 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.

### 9. Beneficios

Principales beneficios, tanto para la Administración como para el ciudadano, obtenidos gracias a la implementación de este algoritmo/sistema de IA.

#### 9.1 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.

##### *[a] Mejora de los servicios públicos*

- ▶ 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

##### *[b] Mejora de la eficiencia administrativa*

- ▶ 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

### [c] Capacidades de gobierno abierto

- ▶ Mayor transparencia de las operaciones del sector público
- ▶ Aumento de la participación pública en las acciones gubernamentales y en la formulación de políticas
- ▶ Mejora del control público y la influencia en las acciones y políticas gubernamentales

## 10. Riesgos

En su opinión, ¿cuáles son los principales riesgos, tanto para la Administración como para el ciudadano, de la implementación de este algoritmo/sistema de IA?

### 10.1 Riesgos

Seleccionar los riesgos que presenta o podría comportar el uso de este algoritmo/sistema de IA. Categorías extraídas del repositorio OASI, de Éticas.

- ▶ 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: ...

## Bloque 4. Supervisión y transparencia

### 11. Intervención humana

¿Qué rol juega la supervisión humana?

- ▶ El resultado del algoritmo es supervisado por una persona que tiene la decisión final
- ▶ El resultado del algoritmo es consultado por una persona, pero esta no puede modificar la decisión

- ▶ El resultado del algoritmo es supervisado/consultado por una persona, pero esta no tiene el conocimiento/capacidad para evaluarlo
- ▶ El resultado no se supervisa/consulta
- ▶ Otro: ...

### 11.1 Intervención humana (detalles)

En caso de responder la tercera opción, explicar por qué no se supervisa. Si ha respondido con una opción alternativa, aportar más detalles sobre el grado de intervención humana.

## 12. Auditoría

¿El algoritmo ha superado un sistema de auditoría? Seleccionar una opción si ha pasado una auditoría:

- ▶ Se realizó o está pasando actualmente una auditoría interna
- ▶ Se realizó o está pasando actualmente una auditoría externa
- ▶ Próximamente pasará una auditoría interna (ya prevista)
- ▶ Próximamente pasará una auditoría externa (ya prevista)
- ▶ Otro ...

## 13. Transparencia

Grado de accesibilidad a la información de este algoritmo. ¿Existe algún documento público y/o página web donde se exponga la siguiente información? (Marque las que sí)

- ▶ 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: ...

## 14. Visibilidad del ciudadano

¿El ciudadano afectado directamente o indirectamente por la decisión/resultado del algoritmo es consciente de que se está usando este algoritmo? 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.

- ▶ 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

## 15. Enlaces y fuentes

¿Existe algún documento público y/o página web donde se incluya información sobre la utilización de este algoritmo?

## 4. RESULTADOS

A continuación, se muestran las fichas de los casos, es decir, toda la información recogida para los **12 sistemas de inteligencia artificial identificados**. La información de cada sistema está sintetizada en una tabla donde cada fila representa uno de los campos explicados anteriormente.

En la siguiente tabla se recogen los nombres de los algoritmos, junto con la administración que los ha puesto en funcionamiento:

Algoritmo	Administración
1 Sistema de identificación de demandas, problemas y respuestas (DPR)	Ayuntamiento de Barcelona e Instituto Municipal de Servicios Sociales (perteneciente al Ayuntamiento)
2 Sistema de apoyo a la clasificación de incidencias - MARIO	
3 Procesador de casos IRIS	
4 Sistemas conversacionales, infoCanal Alzheimer y Fondo 0-16	Ayuntamiento de Barcelona
5 Transcripción de informes de emergencia social (CUESB)	
6 RisCanvi	
7 Motor de reglas de prestaciones sociales	Generalitat de Catalunya - Área de Tecnologías de la Información y las Comunicaciones
8 Recomendador de prestaciones	
9 Identificación necesidades en la intervención social	
10 Xatbots conversacionales	
11 Videoidentificación para la obtención del idCAT Móvil	Administración Abierta de Catalunya (AOC)
12 Servicio de automatización de los informes de pobreza energética	

De los 12 sistemas de IA mapeados, 6 se encuentran en el nivel local y 6 en el nivel autonómico. El único municipio con sistemas de IA es Barcelona. Además, 7 ya están implementados y los restantes, en fase piloto.

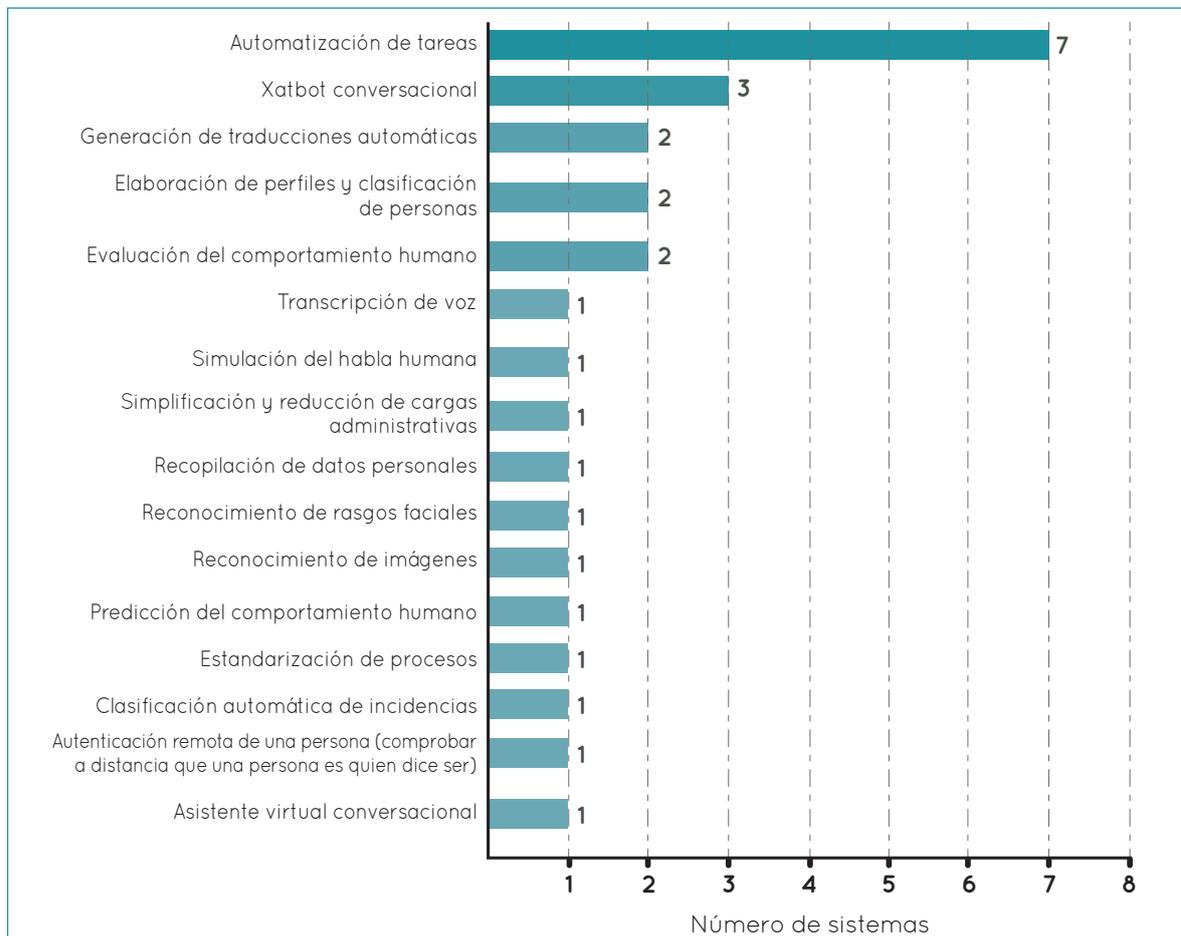
En cuanto al grado de intervención humana, solo 1 sistema no tiene ningún tipo de supervisión humana (sistema MARIO), pero teniendo en cuenta que se limita a recomendar al usuario cómo clasificar su incidencia en una serie de temáticas, es razonable que no se requiera supervisión humana. En el resto de casos,

en función del uso, o bien se supervisan los resultados periódicamente o bien es una persona quien tiene la decisión final. Asimismo, hasta 8 de los sistemas ya han pasado recientemente o pasarán una auditoría, y también en 9 casos existe información pública sobre las principales características de los sistemas.

Por lo tanto, en términos generales, y pese a encontrarnos en una fase muy reciente de desarrollo, implementación y uso de estas tecnologías en el ámbito de la Administración pública catalana, **el grado de seguridad (entendido como el control del sistema mediante auditorías y supervisión, así como por la difusión que se hace de estas herramientas) es considerable, lo que muestra un punto de partida positivo.**

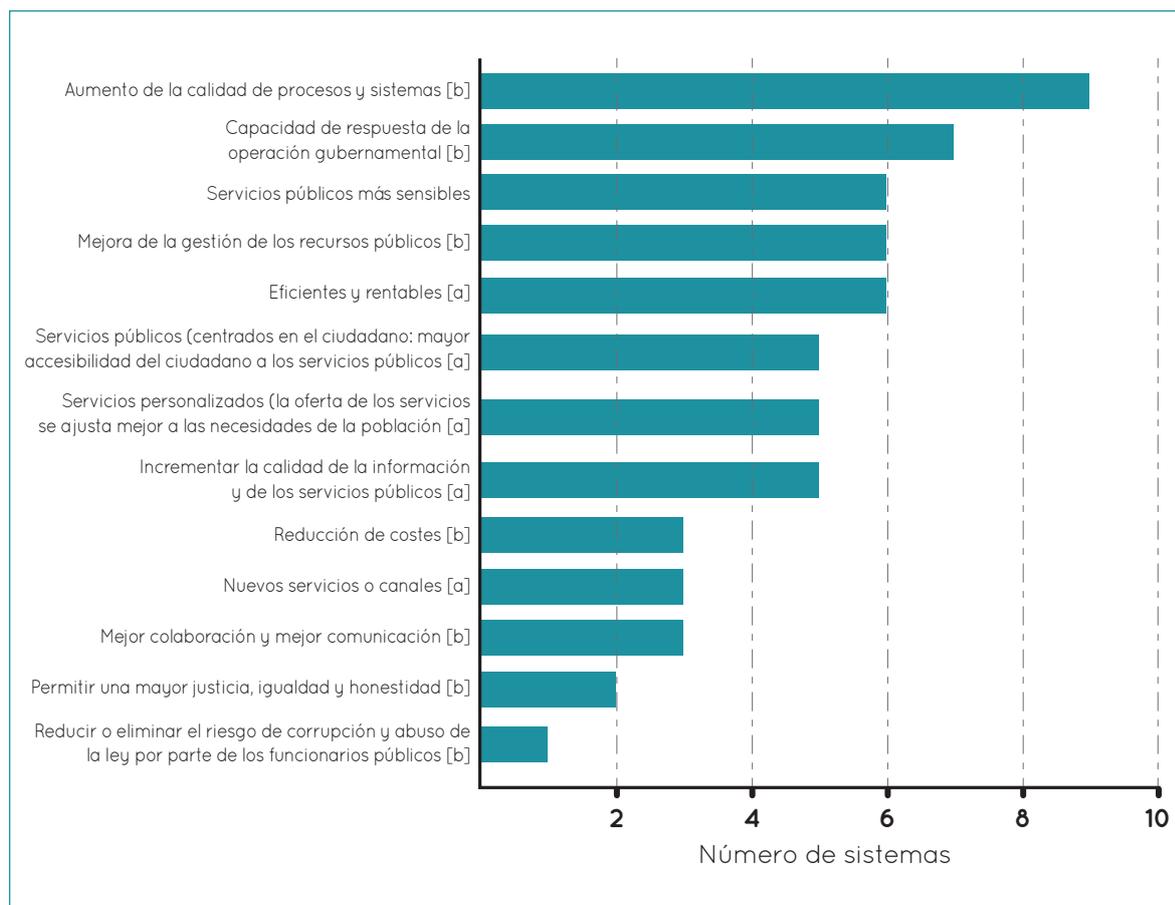
Los objetivos que cumplen estos sistemas de IA, clasificados según los ítems utilizados por el Joint Research Centre (JRC), son distintos. No obstante, destaca que 7 de los 12 sistemas recogidos tienen por objetivo la automatización de tareas. Por lo tanto, la ganancia en eficiencia y liberación del tiempo parece tener un peso importante en la decisión de implementar estos sistemas.

### Gráfico 13. Objetivos que cumplen los sistemas de inteligencia artificial



En cuanto a los beneficios, y de acuerdo también con las categorías extraídas del JRC, la mayor parte de los expertos han destacado el aumento de la calidad de los procesos y sistemas en 9 de los 12 algoritmos estudiados, seguido por la capacidad de respuesta de la operación gubernamental.

### Gráfico 14. Beneficios que aportan los sistemas de inteligencia artificial



Finalmente, los riesgos identificados por los expertos son pocos. En este punto también es importante mencionar que, tal y como se destacó en más de una entrevista, cualquier proceso implementado por las AA. PP. debe ser garante. Eso hace que se reduzca de forma notoria el abanico de sistemas que pueden utilizarse. Tan solo en 3 de los sistemas recogidos se ha considerado que podrían suponer una amenaza a la privacidad, mientras que 2 sistemas podrían comportar un riesgo de difundir información errónea. Las discriminaciones por razón de raza, nivel socioeconómico u otros tipos también han sido identificadas en uno o dos sistemas.

Riesgo identificado por el experto	Número de sistemas de IA que lo cumplen
Amenaza a la privacidad	3
Difundir información errónea	2
Otros tipos de discriminación	2
Discriminación socioeconómica	1
Discriminación racial	1

## Fichas informativas

A continuación, se muestran las fichas informativas de los 12 sistemas identificados.

### 1. Inteligencia colectiva - Sistema de identificación de demandas, problemas y respuestas (DPR automático)

CAMPO	RESPUESTA
<b>1. Nombre</b>	<p>Inteligencia colectiva - Sistema de identificación de demandas, problemas y respuestas (DPR automático)</p>
<b>2. Descripción</b>	<p>El proyecto de DPR automático se enmarca en una iniciativa más amplia de implementación de un sistema de apoyo a las decisiones de los y las profesionales basado en herramientas de inteligencia artificial avanzada, que, recogiendo la experiencia pasada (inteligencia colectiva), realice propuestas de diagnosis y de recursos. Se trata de un proyecto que utiliza procesamiento del lenguaje natural (datos no estructurados) para dar respuesta a la necesidad de agilizar el tiempo en el registro de datos requeridos en el sistema de información utilizado por los profesionales de atención social.</p> <p>El Área de Derechos Sociales del Ayuntamiento de Barcelona atiende una media de cincuenta mil primeras visitas al año. Las personas que acuden a los cuarenta centros de servicios sociales repartidos por la ciudad tienen problemas económicos, de dependencia, por enfermedad mental, de alcoholismo, pueden necesitar ayuda psicológica, de adaptación, pueden sufrir una situación de violencia de género, etc. Estas problemáticas tan diversas son atendidas por una plantilla de más de setecientos profesionales, entre trabajadores/as sociales, psicólogos/as y educadores/as sociales.</p> <p>Cuando la persona llega al centro, se la atiende en unas cabinas privadas. El trabajador/a social graba la conversación y, al acabar, transcribe la problemática, así como la ayuda o servicio al cual se la ha derivado. En el sistema interno, la información se ordena con tres letras: demanda (D), problema (P), recurso (R). Actualmente, el Ayuntamiento dispone de centenares de miles de entrevistas, muchas de las cuales acaban siendo repetitivas porque los problemas son muy parecidos.</p> <p>El DPR automático da apoyo al proceso de registro del DPR: demandas, problemas y respuestas (lo que solicita la persona atendida, los problemas que identifican los y las profesionales y las respuestas que se proponen para afrontarlos).</p> <p>La codificación de este DPR es vivida por los y las profesionales como algo que se hace por el hecho de ser obligatorio, pero que aporta poco valor añadido a nivel de atención. Sin embargo, estos datos son de gran utilidad para la organización (para la extracción de indicadores), aunque también podrían servir para aportar información a los y las profesionales (para ver la evolución del caso, en traspaso de casos...).</p>

CAMPO	RESPUESTA
<p><b>2. Descripción</b></p>	<p>El proyecto se inició con una prueba piloto en 2018 en tres centros de servicios sociales. La prueba fue un éxito, tanto a nivel cualitativo (adecuación de las propuestas ofrecidas por el sistema) como cuantitativo (uso que se hizo): su eficacia y su nivel de concordancia fue muy elevado (entre el 75 % y el 95 % en función de la categoría), y se detectó un mayor volumen de multipropuesta (más de una D, P, R por caso) que por medio del registro manual.</p> <p>Una vez finalizado el periodo de prueba, se decidió mantener la funcionalidad en estos tres centros de la ciudad, a la espera de hacer extensiva la herramienta al resto de los territorios.</p> <p>En octubre de 2020 se produjo un cambio en la codificación de los diccionarios de problemáticas y recursos utilizados por los centros de servicios sociales. Eso implicó un nuevo entrenamiento para el sistema, dado que tenía que hacer nuevas asociaciones. Bajo el uso de los nuevos diccionarios, en junio de 2021 se pidió a los centros que grabasen de forma obligatoria tanto las demandas como las problemáticas y los recursos durante un periodo de tiempo, y se volvió a entrenar el sistema.</p> <p>A raíz de la revaloración positiva de la calidad de acierto de la propuesta automática, en octubre de 2021 se desplegó esta nueva funcionalidad para el conjunto de centros de la ciudad de Barcelona, así como para el servicio de atención a la dependencia y el Equipo de Demandas de Estudio de Infancia y Adolescencia en Riesgo.</p> <p>Durante el año 2022 se llevó a cabo un proceso de internalización del sistema de inteligencia colectiva dentro de la infraestructura corporativa del Ayuntamiento. En este proceso se definieron procesos de medición de la calidad y de reentrenamiento automático y manual, que, además, permiten monitorizar el grado de éxito del algoritmo.</p> <p>Actualmente, el uso de la funcionalidad del DPR automático forma ya parte del día a día del trabajo de los equipos de profesionales, considerándose ya como una herramienta consolidada en el sistema y parte del journey de trabajo de los profesionales de atención.</p>
<p><b>3. Objetivo I</b></p>	<p>El DPR automático es una herramienta integrada dentro del sistema de información de los centros de servicios sociales que sirve para automatizar el proceso de codificación dentro del sistema de las demandas recibidas, las problemáticas detectadas y las propuestas de prescripción de recursos que hacen los profesionales. Por lo tanto, el objetivo principal es servir de apoyo a la decisión del trabajador/a social.</p> <p>El sistema ha sido entrenado con un repositorio de trescientas mil entrevistas con técnicas de aprendizaje automático, básicamente los registros de las conclusiones de las citas realizadas por los profesionales. A partir de las conclusiones de la entrevista introducidas por un profesional, el sistema, que ha sido entrenado previamente, analiza todo su contenido y busca concordancias. A partir de esas concordancias, busca qué habían informado los y las profesionales en situaciones similares y propone demandas, problemas y respuestas.</p> <p>A partir de ahí, el profesional valida las propuestas, pudiendo, después de validar o informar de nuevo la D, pedir al sistema que afine P y R otra vez.</p>

CAMPO	RESPUESTA
	De esta forma, los y las profesionales se ahorran la tarea de buscar entre los diccionarios de demandas, problemas y respuestas definidos, ya que el sistema les hace una propuesta a partir de la interpretación de las conclusiones de las citas y el histórico de información introducida en el sistema.
<b>3. Objetivo I</b>	Este proyecto convierte una acción que hasta el momento se tenía que hacer de forma manual en una tarea automática, requiriendo únicamente la validación de los y las profesionales. Con esta herramienta se reducen las horas destinadas a la gestión por parte de los equipos de profesionales, con el fin de poder aumentar el tiempo dedicado a la atención social, y a la vez se consigue un mayor registro y sistematización de la información
<b>4. Objetivo II</b>	<ul style="list-style-type: none"> <li>▶ Automatización de tareas</li> <li>▶ Evaluación del comportamiento humano</li> <li>▶ Elaboración de perfiles y clasificación de personas</li> </ul>
<b>5. Adm. Responsable</b>	Instituto Municipal de Servicios Sociales - Ayuntamiento de Barcelona Área de Derechos Sociales - Ayuntamiento de Barcelona
<b>6. Entidad desarrolladora</b>	El proyecto estuvo liderado en su fase piloto por la Dirección de Innovación del Área de Derechos Sociales, en colaboración con el Instituto Municipal de Servicios Sociales, y ejecutado por la empresa INNOVA. La fase de interiorización del algoritmo fue gestionada por el Instituto Municipal de Informática, en colaboración con la empresa Accenture.
<b>7. Estatus</b>	Implementado: el sistema está implementado y en uso.
<b>8. Fecha de inicio</b>	2018 en formato piloto, 2019 en el Área de Derechos Sociales y octubre de 2021 de forma definitiva para todos los centros de Barcelona
<b>9. Fecha de finalización</b>	
<b>10. Colectivo</b>	Trabajadores/as sociales
<b>11. Beneficios I</b>	<ol style="list-style-type: none"> <li>1) Ayudar a los trabajadores/as sociales a categorizar las entrevistas con los ciudadanos.</li> <li>2) Sistematización de los servicios/prestaciones prescritos a partir de una visión neutral y objetiva de las demandas. El sistema ofrece ayuda guiando al profesional con propuestas basadas en la experiencia anterior (inteligencia colectiva), de forma que el DPR aporta un valor añadido al proceso de atención.</li> <li>3) Eficiencia de recursos destinados: facilita el proceso de registro de DPR, reduce el tiempo destinado a esta tarea y facilita el multirregistro.</li> </ol>

CAMPO	RESPUESTA
<b>12. Beneficios II</b>	<ul style="list-style-type: none"> <li>▶ Incrementar la calidad de la información y de los servicios públicos [a]</li> <li>▶ Servicios personalizados (la oferta de los servicios se ajusta mejor a las necesidades de la población) [a]</li> <li>▶ Servicios públicos (centrados en el ciudadano: mayor accesibilidad del ciudadano a los servicios públicos) [a]</li> <li>▶ Servicios públicos más sensibles, eficientes y rentables [a]</li> <li>▶ Mejora de la gestión de los recursos públicos [b]</li> <li>▶ Aumento de la calidad de procesos y sistemas [b]</li> <li>▶ Capacidad de respuesta de la operación gubernamental [b]</li> </ul>
<b>13. Riesgos I</b>	<p>Puede haber algún sesgo, ya que los trabajadores/as sociales son humanos que tienen su propio sesgo, pero se minimiza gracias a un gran volumen de entrevistas y a su poco impacto en las personas.</p> <p>Las entrevistas forman parte del sistema corporativo municipal/estándares de ciberseguridad. El modelo está alojado en un servidor externo, pero no los datos.</p> <p>Dado que el sistema entra en juego en una 2.ª fase -después de que ya se ha dado una respuesta-, el riesgo puede asociarse a la mala gestión del servicio/proceso, no tanto al sistema algorítmico. El proceso en sí y la forma como se utiliza el sistema ya mitigan los riesgos.</p>
<b>14. Riesgos II</b>	
<b>15. Intervención humana</b>	El resultado del algoritmo es supervisado por una persona que tiene la decisión final.
<b>16. Intervención humana (detalles)</b>	<p>El sistema ofrece una propuesta al profesional de demandas, problemas y respuestas.</p> <p>A partir de ahí, el profesional puede validar la propuesta o informar de nuevo la D y pedir al sistema que afine P y R otra vez. La decisión está, por lo tanto, siempre en manos del profesional.</p>
<b>17. Auditoría</b>	Se realizó o está pasando actualmente una auditoría externa.
<b>18. Transparencia</b>	<ul style="list-style-type: none"> <li>▶ Problemática a la cual se pretende dar solución</li> <li>▶ Objetivos y población que interactúa con el algoritmo</li> <li>▶ Cómo se está implementando</li> <li>▶ Resultados obtenidos hasta ahora</li> </ul>
<b>19. Visibilidad</b>	La información es pública pero poco accesible (depende de la proactividad del ciudadano para informarse).
<b>20. Enlaces</b>	<p><a href="https://ajuntament.barcelona.cat/mesames/noticia/posem-en-marxa-una-prova-pilot-dintel%C2%B7ligencia-col%C2%B7lectiva-a-tres-centres-de-servis-socials/">https://ajuntament.barcelona.cat/mesames/noticia/posem-en-marxa-una-prova-pilot-dintel%C2%B7ligencia-col%C2%B7lectiva-a-tres-centres-de-servis-socials/</a></p> <p><a href="https://ajuntament.barcelona.cat/dretsocials/es/innovacion-social/inteligencia-colectiva">https://ajuntament.barcelona.cat/dretsocials/es/innovacion-social/inteligencia-colectiva</a></p> <p><a href="https://ajuntament.barcelona.cat/premsa/wp-content/uploads/2021/02/Mesura-de-Govern-dInnovacio-Social.pdf">https://ajuntament.barcelona.cat/premsa/wp-content/uploads/2021/02/Mesura-de-Govern-dInnovacio-Social.pdf</a></p>

## 2. Procesador de casos IRIS

CAMPO	RESPUESTA
<b>1. Nombre</b>	Procesador de casos IRIS
<b>2. Descripción</b>	Herramienta de gestión interna que permite acelerar el proceso de categorización en temáticas de las quejas, incidencias y sugerencias que llegan al Área de Derechos Sociales a través de la plataforma RS. El sistema también ofrece tendencias de cuáles son las temáticas más recurrentes según diversos criterios.
<b>3. Objetivo I</b>	El objetivo es mejorar la extracción de datos relevantes a partir de las consultas/quejas que llegan a través del sistema RS, bien sea utilizando sistemas de clasificación automatizados o creando consultas con lenguaje natural.
<b>4. Objetivo II</b>	Automatización de tareas.
<b>5. Adm. Responsable</b>	Ayuntamiento de Barcelona - Área de Derechos Sociales
<b>6. Entidad desarrolladora</b>	Secretaría Técnica de Innovación Social - Área de Derechos Sociales
<b>7. Estatus</b>	Piloto: el sistema está siendo probado.
<b>8. Fecha de inicio</b>	2022
<b>9. Fecha de finalización</b>	En curso
<b>10. Colectivo</b>	Gestores del Área de Derechos Sociales
<b>11. Beneficios I</b>	Mejorar en la eficiencia, dar respuestas más concretas y tener un conocimiento más profundo en la organización de las consultas/quejas que llegan.
<b>12. Beneficios II</b>	<ul style="list-style-type: none"> <li>▶ Servicios públicos más sensibles, eficientes y rentables [a]</li> <li>▶ Capacidad de respuesta de la operación gubernamental [b]</li> <li>▶ Aumento de la calidad de procesos y sistemas [b]</li> </ul>
<b>13. Riesgos I</b>	No existen riesgos derivados, ya que el sistema de clasificación es de uso interno y para facilitar la explotación de datos.
<b>14. Riesgos II</b>	
<b>15. Intervención humana</b>	El resultado del algoritmo es supervisado por una persona que tiene la decisión final.
<b>16. Intervención humana (detalles)</b>	La aplicación hace una sugerencia de temas tratados en la queja, incidencia o sugerencia enviada por el ciudadano. Pueden ser editadas o rechazadas.
<b>17. Auditoría</b>	Se realizó o está pasando actualmente una auditoría interna.

CAMPO	RESPUESTA
18. Transparencia	
19. Visibilitat	No se informa.
20. Enlaces	De momento no existe ninguna información pública, ya que está en fase piloto.

### 3. Sistema de apoyo a la clasificación de incidencias - MARIO

CAMPO	RESPUESTA
1. Nombre	Sistema de apoyo a la clasificación de incidencias – MARIO
2. Descripción	<p>IRIS es el servicio que permite a la ciudadanía comunicar incidencias o enviar reclamaciones al Ayuntamiento de Barcelona para que las resuelva. A través de IRIS, la ciudadanía puede hacer llegar informaciones y consultas, pero también quejas y sugerencias de mejora. En este proceso, el ciudadano o ciudadana que avisa de la incidencia debe clasificarla utilizando un árbol de temáticas que se le ofrece a través de la aplicación informática. Esta clasificación es importante, porque se utiliza para dirigir la incidencia directamente al departamento responsable, agilizando así el proceso de respuesta. Los errores en la clasificación temática ocasionan respuestas inadecuadas y retrasos en la resolución de las incidencias, afectando así a la calidad del servicio ofrecido. Dentro del proyecto de actualización del servicio IRIS, se ha desarrollado un módulo llamado MARIO, basado en algoritmos de aprendizaje automático (machine learning) y procesamiento del lenguaje natural (una de las tecnologías enmarcadas en la A), para simplificar al ciudadano o ciudadana el proceso de clasificación de las incidencias. A partir del análisis del texto libre que describe la incidencia, MARIO sugiere al ciudadano o ciudadana las categorías más probables donde la incidencia tiene cabida para que elija la categoría más adecuada.</p>
3. Objetivo I	Minimizar la tasa de error en la clasificación inicial de las incidencias y así reducir los procesos de clasificación manual.
4. Objetivo II	Clasificación automática de incidencias.
5. Adm. Responsable	Dirección de servicios de información y atención ciudadana - Ayuntamiento de Barcelona
6. Entidad desarrolladora	
7. Estatus	Implementado: el sistema está implementado y en uso.
8. Fecha de inicio	Abril 2021

CAMPO	RESPUESTA
9. Fecha de finalización	
10. Colectivo	Ciudadanía con acceso a medios digitales
11. Beneficios I	Ahorrar la tarea manual del personal funcionario reclasificando incidencias y agilizar la respuesta de la Administración.
12. Beneficios II	Incrementar la calidad de la información y de los servicios públicos [a] Aumento de la calidad de procesos y sistemas [b]
13. Riesgos I	
14. Riesgos II	
15. Intervención humana	El resultado no se supervisa/consulta.
16. Intervención humana (detalles)	No se supervisa, dado que, en principio, el algoritmo ya orienta al ciudadano hacia una clasificación adecuada para la incidencia que está tramitando.
17. Auditoría	
18. Transparencia	Problemática a la cual se pretende dar solución
19. Visibilitat	La información es pública y accesible (comunicación pasiva).
20. Enlaces	<a href="https://media-edg.barcelona.cat/wp-content/uploads/2023/08/03111619/drets-digitals.pdf">https://media-edg.barcelona.cat/wp-content/uploads/2023/08/03111619/drets-digitals.pdf</a>

## 4. Sistemas conversacionales, infoCanal Alzheimer y Fondo 0-16

CAMPO	RESPUESTA
1. Nombre	Sistemas conversacionales, infoCanal Alzheimer y Fondo 0-16
2. Descripción	Es un canal de información nuevo presente en las webs municipales para dar información a la ciudadanía sobre el Alzheimer y sobre el Fondo 0-16. En ambos casos, las respuestas de los sistemas conversacionales se construyen a partir de fuentes validadas y propias.
3. Objetivo I	Ofrecer a las personas información seleccionada a través de conversaciones en lenguaje natural.
4. Objetivo II	Xatbot conversacional.
5. Adm. Responsable	Ayuntamiento de Barcelona - Área de Derechos Sociales

CAMPO	RESPUESTA
<b>6. Entidad desarrolladora</b>	Secretaría Técnica de Innovación Social - Área de Derechos Sociales
<b>7. Estatus</b>	Implementado: el sistema está implementado y en uso.
<b>8. Fecha de inicio</b>	2022 infoCanal y 2023 Fondo 0-16
<b>9. Fecha de finalización</b>	En curso
<b>10. Colectivo</b>	Ciudadanía en general
<b>11. Beneficios I</b>	Ampliar los canales de información y poder ofrecer respuestas de forma más fácil.
<b>12. Beneficios II</b>	Incrementar la calidad de la información y de los servicios públicos [a]
<b>13. Riesgos I</b>	El principal riesgo es que el sistema conversacional no dé respuestas adecuadas.
<b>14. Riesgos II</b>	Difundir información errónea
<b>15. Intervención humana</b>	Todas las interacciones de las personas y las respuestas del sistema conversacional se registran y pueden ser evaluadas.
<b>16. Intervención humana (detalles)</b>	
<b>17. Auditoría</b>	Se realizó o está pasando actualmente una auditoría interna.
<b>18. Transparencia</b>	Se informa en los entornos donde están implementados.
<b>19. Visibilitat</b>	La información es pública y accesible (comunicación pasiva).
<b>20. Enlaces</b>	InfoCanal Alzheimer: <a href="https://ajuntament.barcelona.cat/sanitalisalut/ca/canal/barcelona-lalzheimer">https://ajuntament.barcelona.cat/sanitalisalut/ca/canal/barcelona-lalzheimer</a> Fons 0-16: <a href="https://ajuntament.barcelona.cat/serveissocials/ca/canal/ajuts-durgencia-social-families-amb-infants-de-0-16-anys">https://ajuntament.barcelona.cat/serveissocials/ca/canal/ajuts-durgencia-social-families-amb-infants-de-0-16-anys</a>

## 1. Transcripción de informes de emergencia social (CUESB)

CAMPO	RESPUESTA
<b>1. Nombre</b>	Transcripción de informes de emergencia social (CUESB)
<b>2. Descripción</b>	<p>En la búsqueda de soluciones para agilizar tareas de bajo valor añadido por parte de las profesionales del Área de Derechos Sociales, se identifican los procesos de transcripción automatizados mediante IA como un elemento relevante. Los servicios de emergencia del CUESB realizan informes después de cada intervención, lo que a día de hoy obliga al personal a volver a la oficina y rellenar formularios no sistematizados.</p> <p>Con este piloto se busca que las profesionales puedan dictar el informe <i>in situ</i> o en cualquier momento de la intervención y que este se transcriba.</p>
<b>3. Objetivo I</b>	Agilizar el proceso de creación de los informes a través de sistemas de transcripción.
<b>4. Objetivo II</b>	Transcripción de voz.
<b>5. Adm. Responsable</b>	Ayuntamiento de Barcelona - Área de Derechos Sociales
<b>6. Entidad desarrolladora</b>	Secretaría Técnica de Innovación Social - Área de Derechos Sociales
<b>7. Estatus</b>	En desarrollo: el sistema se está desarrollando.
<b>8. Fecha de inicio</b>	2024
<b>9. Fecha de finalización</b>	El piloto se ejecutará durante 6 meses y se evaluará
<b>10. Colectivo</b>	Trabajadoras del servicio de emergencias sociales del Ayuntamiento de Barcelona
<b>11. Beneficios I</b>	Reducción de tiempo de elaboración de los informes. Agilización de tareas de bajo valor añadido.
<b>12. Beneficios II</b>	Aumento de la calidad de procesos y sistemas [b]
<b>13. Riesgos I</b>	El sistema no tiene riesgos asociados, ya que la transcripción se ofrece al usuario como complemento y es revisable y editable.
<b>14. Riesgos II</b>	
<b>15. Intervención humana</b>	El resultado del algoritmo es supervisado por una persona que tiene la decisión final.
<b>16. Intervención humana (detalles)</b>	
<b>17. Auditoría</b>	Próximamente pasará una auditoría interna (ya prevista).
<b>18. Transparencia</b>	Para la transcripción se utiliza un servicio externo. La información es pública en su página web.

CAMPO	RESPUESTA
19. Visibilitat	No se informa.
20. Enlaces	Servicio externo: <a href="https://www.speechmatics.com/">https://www.speechmatics.com/</a>

## 6. RisCanvi

CAMPO	RESPUESTA
1. Nombre	RisCanvi
2. Descripción	<p>El proyecto RisCanvi de evaluación y valoración del riesgo tiene sus orígenes fundacionales a finales del año 2007, a partir de las recomendaciones de la Comisión para el estudio de las medidas de prevención de la reincidencia en delitos graves (Resolución JUS/2363/200, DOGC n. 4937, de 24 de julio de 2007). El primer punto de las recomendaciones mencionaba «() el establecimiento de un procedimiento técnico específico de valoración del grado de riesgo que sea compartido por todas las instituciones competentes –administración de justicia, servicios penitenciarios y cuerpos policiales».</p> <p>La evaluación RisCanvi es universal y continua: debe evaluarse el riesgo de todos los internos e internas de todos los centros penitenciarios y en libertad condicional, y esa evaluación debe actualizarse periódicamente. Como norma general, las evaluaciones tienen una vigencia máxima de 6 meses.</p> <p>Las escalas RisCanvi (RisCanvi-Screening y RisCanvi-Completa) han sido diseñadas para predecir el riesgo de aparición de cinco tipos de comportamientos: violencia autodirigida, violencia intra-institucional, reincidencia general, reincidencia violenta y quebrantamiento de condena.</p> <p>Desde su implementación en 2008, se han evaluado un total de 30.832 internos e internas diferentes, con un total de 116.680 escalas evaluadas. Actualmente se utiliza la tercera versión del algoritmo: RisCanvi v3.</p>

CAMPO	RESPUESTA
<p><b>3. Objetivo I</b></p>	<p>Los objetivos del protocolo de valoración RisCanvi se dividen en objetivos generales y específicos</p> <p><b>Generales:</b></p> <ol style="list-style-type: none"> <li>1. Mejorar las predicciones individualizadas del riesgo de violencia futura.</li> <li>2. Generalizar la aplicación de herramientas de valoración del riesgo como procedimiento de trabajo de los y las profesionales penitenciarios/as.</li> <li>3. Introducir la gestión del riesgo como herramienta de intervención.</li> <li>4. Potenciar la coordinación institucional en la detección y seguimiento de los casos.</li> </ol> <p><b>Específicos:</b></p> <ol style="list-style-type: none"> <li>1. Hacer predicciones dinámicas ajustadas a los cambios de los internos e internas.</li> <li>2. Identificar internos e internas en riesgo, con independencia del delito cometido.</li> <li>3. Introducir técnicas de cribado y detección de un riesgo de violencia de forma rápida y sencilla.</li> <li>4. Generalizar la evaluación del riesgo en toda la población penitenciaria.</li> <li>5. Evaluar el efecto de la intervención con programas específicos, en relación con la gestión del riesgo.</li> <li>6. Homogeneizar criterios entre los/las profesionales y cualificar su capacitación técnica.</li> <li>7. Organizar procedimientos sistemáticos de valoración de riesgos específicos.</li> <li>8. Realizar análisis retrospectivos y de calidad de las decisiones ya tomadas, para mejorar el proceso de valoración.</li> </ol>
<p><b>4. Objetivo II</b></p>	<ul style="list-style-type: none"> <li>▸ Evaluación del comportamiento humano</li> <li>▸ Predicción del comportamiento humano</li> <li>▸ Elaboración de perfiles y clasificación de personas</li> </ul>
<p><b>5. Adm. Responsable</b></p>	<p>Secretaría de Medidas Penales, Reinserción y Atención a la Víctima. Departamento de Justicia de la Generalitat de Cataluña</p>
<p><b>6. Entidad desarrolladora</b></p>	
<p><b>7. Estatus</b></p>	<p>Implementado: el sistema está implementado y en uso.</p>
<p><b>8. Fecha de inicio</b></p>	<p>2008</p>

CAMPO	RESPUESTA
9. Fecha de finalización	
10. Colectivo	Población reclusa
11. Beneficios I	
12. Beneficios II	<ul style="list-style-type: none"> <li>▶ Servicios públicos más sensibles, eficientes y rentables [a]</li> <li>▶ Mejora de la gestión de los recursos públicos [b]</li> </ul>
13. Riesgos I	
14. Riesgos II	<ul style="list-style-type: none"> <li>▶ Discriminación socioeconómica</li> <li>▶ Discriminación racial</li> <li>▶ Otros tipos de discriminación</li> </ul>
15. Intervención humana	El resultado del algoritmo es supervisado por una persona que tiene la decisión final.
16. Intervención humana (detalles)	El algoritmo asigna un nivel de riesgo que puede ser corregido por la persona.
17. Auditoría	Se realizó o está pasando actualmente una auditoría interna.
18. Transparencia	<ul style="list-style-type: none"> <li>▶ Problemática a la cual se pretende dar solución</li> <li>▶ Cómo se está implementando</li> <li>▶ Objetivos y población que interactúa con el algoritmo</li> <li>▶ Resultados obtenidos hasta ahora</li> </ul>
19. Visibilitat	La información es pública pero poco accesible (depende de la proactividad del ciudadano para informarse).
20. Enlaces	<p><a href="https://justicia.gencat.cat/web/.content/home/ambits/reinsercio_i_serveis_peni/manual-aplicacio-protocol-avaluacio-riscanvi.pdf">https://justicia.gencat.cat/web/.content/home/ambits/reinsercio_i_serveis_peni/manual-aplicacio-protocol-avaluacio-riscanvi.pdf</a></p> <p><a href="https://digitalfuturesociety.com/es/podcasts/capitulo-1-riscanvi-i-el-algoritmo-de-la-carcel/">https://digitalfuturesociety.com/es/podcasts/capitulo-1-riscanvi-i-el-algoritmo-de-la-carcel/</a></p> <p><a href="https://eapc.gencat.cat/web/.content/home/biblioteca/fons-coleccions/docs_eapc/tic/2018/03_bones_practiques_adm_electronica/10179_eRis-Canvi-EAP-marc-2018-PRE-1.pdf">https://eapc.gencat.cat/web/.content/home/biblioteca/fons-coleccions/docs_eapc/tic/2018/03_bones_practiques_adm_electronica/10179_eRis-Canvi-EAP-marc-2018-PRE-1.pdf</a></p>

## 7. Motor de reglas de prestaciones sociales

CAMPO	RESPUESTA
<b>1. Nombre</b>	Motor de reglas de prestaciones sociales
<b>2. Descripción</b>	El motor de reglas es una de las piezas de la nueva plataforma eSocial (nueva plataforma tecnológica para dar respuesta a las necesidades de gestión del Departamento de Derechos Sociales), y está pensada para aplicar los criterios por los cuales se otorga una prestación social durante su tramitación.
<b>3. Objetivo I</b>	Verificar si un ciudadano o entidad cumple los criterios para recibir una prestación social. Actualmente, existen cinco algoritmos desarrollados: necesidades básicas, gastos del hogar, PNC jubilación, complemento de la PNC jubilación y prestación por nacimiento, acogida y adopción (familias).
<b>4. Objetivo II</b>	Automatización de tareas.
<b>5. Adm. Responsable</b>	Generalitat de Catalunya - Departamento de Derechos Sociales
<b>6. Entidad desarrolladora</b>	Centro de Telecomunicaciones y Tecnologías de la Información (CTT)
<b>7. Estatus</b>	Implementado: el sistema está implementado y en uso.
<b>8. Fecha de inicio</b>	Diciembre de 2018
<b>9. Fecha de finalización</b>	
<b>10. Colectivo</b>	Están desarrollados cinco algoritmos: necesidades básicas, gastos del hogar, PNC jubilación, complemento de la PNC jubilación y prestación por nacimiento, acogida y adopción (familias). Y se benefician todos los ciudadanos que necesitan esta prestación, es decir, personas mayores, familias o ciudadanos en situación de exclusión social.
<b>11. Beneficios I</b>	Eficiencia en la tramitación. Reducción de los plazos de tramitación.
<b>12. Beneficios II</b>	<ul style="list-style-type: none"> <li>▶ Servicios públicos (centrados en el ciudadano: mayor accesibilidad del ciudadano a los servicios públicos) [a]</li> <li>▶ Capacidad de respuesta de la operación gubernamental [b]</li> <li>▶ Mejora de la gestión de los recursos públicos [b]</li> <li>▶ Aumento de la calidad de procesos y sistemas [b]</li> </ul>
<b>13. Riesgos I</b>	No se han detectado.
<b>14. Riesgos II</b>	
<b>15. Intervención humana</b>	Se supervisa el buen funcionamiento del algoritmo de forma periódica.

CAMPO	RESPUESTA
<b>16. Intervención humana (detalles)</b>	
<b>17. Auditoría</b>	Próximamente pasará una auditoría interna (ya prevista).
<b>18. Transparencia</b>	<ul style="list-style-type: none"> <li>▶ Tipo de algoritmo utilizado</li> <li>▶ Problemática a la cual se pretende dar solución</li> <li>▶ Cómo se está implementando, objetivos y población que interactúa con el algoritmo</li> <li>▶ Resultados obtenidos hasta ahora</li> </ul>
<b>19. Visibilidad</b>	
<b>20. Enlaces</b>	<a href="https://ctti.gencat.cat/ca/detalls/detallnoticia/Pla_Transformacio_Digital_Drets_Socials_noticia_2023">https://ctti.gencat.cat/ca/detalls/detallnoticia/Pla_Transformacio_Digital_Drets_Socials_noticia_2023</a>

## 8. Recomendador de prestaciones

CAMPO	RESPUESTA
<b>1. Nombre</b>	Recomendador de prestaciones
<b>2. Descripción</b>	Facilitar el acceso a la información de prestaciones sociales.
<b>3. Objetivo I</b>	Implementar un xatbot que ayude a la ciudadanía a conocer qué tipo de ayuda puede recibir de la Administración.
<b>4. Objetivo II</b>	<ul style="list-style-type: none"> <li>▶ Simulación del habla humana</li> <li>▶ Generación de traducciones automáticas</li> <li>▶ Xatbot conversacional</li> </ul>
<b>5. Adm. Responsable</b>	Generalitat de Catalunya - Departamento de Derechos Sociales
<b>6. Entidad desarrolladora</b>	Centro de Telecomunicaciones y Tecnologías de la Información (CTT)
<b>7. Estatus</b>	Piloto: el sistema está siendo probado.
<b>8. Fecha de inicio</b>	Julio 2023
<b>9. Fecha de finalización</b>	
<b>10. Colectivo</b>	Ciudadanía en general
<b>11. Beneficios I</b>	Para la Administración, crear otro canal de comunicación con la ciudadanía.

CAMPO	RESPUESTA
12. Beneficios II	<ul style="list-style-type: none"> <li>▶ Servicios personalizados (la oferta de los servicios se ajusta mejor a las necesidades de la población) [a]</li> <li>▶ Incrementar la calidad de la información y de los servicios públicos [a]</li> <li>▶ Nuevos servicios o canales [a]</li> <li>▶ Capacidad de respuesta de la operación gubernamental [b]</li> </ul>
13. Riesgos I	Compartición de datos personales.
14. Riesgos II	
15. Intervención humana	El resultado del algoritmo es supervisado periódicamente.
16. Intervención humana (detalles)	
17. Auditoría	Próximamente pasará una auditoría interna (ya prevista).
18. Transparencia	
19. Visibilitat	
20. Enlaces	

## 9. Identificación necesidades en la intervención social

CAMPO	RESPUESTA
1. Nombre	Identificación necesidades en la intervención social
2. Descripción	Durante la intervención social, el empleado público, con el consentimiento del ciudadano, puede activar una funcionalidad de la herramienta de intervención social eSocial que, a través de la transcripción de la parte esencial de la conversación con el ciudadano, identifica de forma automática las situaciones de necesidad.
3. Objetivo I	Facilitar el trabajo del trabajador social.
4. Objetivo II	<ul style="list-style-type: none"> <li>▶ Automatización de tareas</li> <li>▶ Recopilación de datos personales</li> <li>▶ Generación de traducciones automáticas</li> </ul>
5. Adm. Responsable	Generalitat de Catalunya - Departamento de Derechos Sociales
6. Entidad desarrolladora	Centro de Telecomunicaciones y Tecnologías de la Información (CTT)

CAMPO	RESPUESTA
<b>7. Estatus</b>	Piloto: el sistema está siendo probado.
<b>8. Fecha de inicio</b>	Diciembre de 2023
<b>9. Fecha de finalización</b>	
<b>10. Colectivo</b>	Ciudadanía en general
<b>11. Beneficios I</b>	Eficiencia administrativa. Homogeneización del criterio de la Administración.
<b>12. Beneficios II</b>	<ul style="list-style-type: none"> <li>▶ Servicios personalizados (la oferta de los servicios se ajusta mejor a las necesidades de la población) [a]</li> <li>▶ Servicios públicos (centrados en el ciudadano: mayor accesibilidad del ciudadano a los servicios públicos) [a]</li> <li>▶ Capacidad de respuesta de la operación gubernamental [b]</li> <li>▶ Aumento de la calidad de procesos y sistemas [b]</li> <li>▶ Mejor colaboración y mejor comunicación [b]</li> </ul>
<b>13. Riesgos I</b>	
<b>14. Riesgos II</b>	
<b>15. Intervención humana</b>	El resultado del algoritmo es supervisado por una persona que tiene la decisión final.
<b>16. Intervención humana (detalles)</b>	
<b>17. Auditoría</b>	
<b>18. Transparencia</b>	
<b>19. Visibilidad</b>	Se informa directamente al ciudadano (comunicación activa).
<b>20. Enlaces</b>	

## 10. Xatbots conversacionales de la Administración Abierta de Cataluña (AOC)

CAMPO	RESPUESTA
1. Nombre	<p>Xatbots conversacionales de la Administración Abierta de Cataluña (AOC)</p>
2. Descripción	<p><b>Problemática detectada:</b></p> <p>Necesidad de asistencia constante: los empleados públicos y la ciudadanía requerían apoyo e información continua las 24 horas del día, los 7 días de la semana, pero la atención presencial o telefónica no podía ofrecer esta disponibilidad total.</p> <p>Aumento en las consultas e incidencias: con el crecimiento de los trámites digitales y el uso del portal de transparencia, había un aumento en la demanda de información y asistencia.</p> <p><b>Solución por la que se optó e implementación:</b></p> <p>Desarrollo de asistentes conversacionales: para abordar esta problemática, se optó por crear xatbots con capacidad para interpretar y responder preguntas mediante el uso de algoritmos de IA.</p> <p>Entrenamiento con datos existentes: los algoritmos fueron entrenados con información ya existente, como las preguntas frecuentes del portal de apoyo de la AOC, estructuradas en preguntas-respuestas y flujos guiados de conversación.</p> <p>Implementación en diversas áreas específicas: los xatbots fueron integrados en diferentes áreas relacionadas con el proceso de identificación digital, obtención, uso y renovación del idCAT Certificado y navegación por el portal de transparencia y e-TRAM 2.0.</p> <p><b>Resultados obtenidos hasta ahora :</b></p> <p>Mejora en la experiencia del usuario: los usuarios pueden acceder a información y resolver dudas de forma más ágil e inmediata, reduciendo la dependencia del apoyo presencial o telefónico.</p> <p>Incremento de la autorresolución: los xatbots han permitido a los usuarios resolver consultas e incidencias sin la necesidad de interacciones humanas, incrementando la eficiencia y liberando recursos.</p> <p>Supervisión y mejora continua: el análisis de los registros de conversación y los feedbacks de los usuarios han contribuido a mejorar los algoritmos, añadir nuevos flujos de conversación y ampliar el contenido para adaptarse mejor a las necesidades de los usuarios.</p>

CAMPO	RESPUESTA
<b>3. Objetivo I</b>	<p>El sistema tiene como principal objetivo facilitar los procesos de identificación digital y el acceso a la información del portal de transparencia, mejorar la experiencia de los usuarios e incrementar la autorresolución de las incidencias y consultas de los ciudadanos y empleados públicos. El sistema está desarrollado para tres funciones específicas:</p> <ol style="list-style-type: none"> <li>1. Asistencia en identificación: se da apoyo a los usuarios en los procesos de identificación y validación de la identidad necesarios para realizar un trámite digital.</li> <li>2. Apoyo a la solicitud y gestión de ciertos documentos: ayuda para resolver las dudas y preguntas más comunes relacionadas con la obtención, uso y renovación del idCAT Certificado.</li> <li>3. Aplicación de asistentes conversacionales por áreas específicas, como el portal de transparencia y e-TRAM 2.0, proporcionando información específica y facilitando la navegación en estas áreas.</li> </ol>
<b>4. Objetivo II</b>	<ul style="list-style-type: none"> <li>▶ Automatización de tareas</li> <li>▶ Xatbot conversacional</li> <li>▶ Asistente virtual conversacional</li> </ul>
<b>5. Adm. Responsable</b>	Administración Abierta de Cataluña (AOC)
<b>6. Entidad desarrolladora</b>	ONE MILLION BOT, S.L.
<b>7. Estatus</b>	Implementado: el sistema está implementado y en uso.
<b>8. Fecha de inicio</b>	21 de enero de 2021
<b>9. Fecha de finalización</b>	
<b>10. Colectivo</b>	Es un servicio de apoyo a empleados públicos y ciudadanía
<b>11. Beneficios I</b>	<p><b>Beneficios para la Administración:</b></p> <p>Optimización de recursos: reducción de carga de trabajo para los empleados públicos, ya que los xatbots gestionan consultas frecuentes y repetitivas, liberándolos para tareas más complejas y especializadas.</p> <p>Mejora de la eficiencia: aumento de la eficiencia en la gestión de información y resolución de consultas, permitiendo una respuesta inmediata y precisa a un gran volumen de consultas simultáneas.</p> <p>Análisis de datos: el análisis de los registros de conversaciones ofrece una valiosa fuente de datos para la Administración, permitiendo identificar patrones, áreas de mejora y necesidades de los usuarios para adaptar mejor los servicios futuros.</p> <p>Reducción de costes: reducción de los costes asociados con la atención al cliente presencial o telefónica, ya que los xatbots proporcionan asistencia digital sin necesidad de recursos humanos. .</p>

CAMPO	RESPUESTA
<p><b>11. Beneficios I</b></p>	<p><b>Beneficios para el ciudadano:</b></p> <p>Acceso 24x7 a la información: disponibilidad constante para acceder a información y asistencia sin restricciones de horas o días, mejorando la comodidad y la experiencia del usuario.</p> <p>Respuesta inmediata: obtención de respuestas rápidas e inmediatas a consultas y dudas, evitando tiempos de espera y agilizando los trámites o procesos que requieren información específica.</p> <p>Autorresolución de consultas: capacidad de autorresolver consultas o incidencias sin la necesidad de recurrir a personal de atención a los usuarios (CAU), dando a los ciudadanos más autonomía en su proceso de interacción con la Administración.</p> <p>Mejora de la experiencia del usuario: a través de la navegación guiada y la respuesta precisa a preguntas específicas, se ha mejorado la experiencia de los usuarios en la interacción con los servicios públicos, facilitando la comprensión y la resolución de dudas.</p>
<p><b>12. Beneficios II</b></p>	<ul style="list-style-type: none"> <li>▶ Servicios personalizados (la oferta de los servicios se ajusta mejor a las necesidades de la población) [a]</li> <li>▶ Servicios públicos (centrados en el ciudadano: mayor accesibilidad del ciudadano a los servicios públicos) [a]</li> <li>▶ Incrementar la calidad de la información y de los servicios públicos [a]</li> <li>▶ Servicios públicos más sensibles, eficientes y rentables [a]</li> <li>▶ Nuevos servicios o canales [a]</li> <li>▶ Reducción de costes [b]</li> <li>▶ Mejora de la gestión de los recursos públicos [b]</li> <li>▶ Aumento de la calidad de procesos y sistemas [b]</li> <li>▶ Mejor colaboración y mejor comunicación [b]</li> </ul>
<p><b>13. Riesgos I</b></p>	<p><b>Riesgos para la Administración:</b></p> <p>Dependencia tecnológica e interrupciones en el servicio: si los xatbots cometen errores o hay interrupciones en el funcionamiento del sistema, eso puede provocar una interrupción en la atención a los usuarios y generar desconfianza en la fiabilidad del servicio.</p> <p>Seguridad de los datos y privacidad: pese a las medidas de seguridad tomadas, existe el riesgo de vulnerabilidades en la protección de datos, especialmente por el registro de conversaciones, que podrían poner en peligro la privacidad de los usuarios (eso puede suceder si el usuario introduce datos personales innecesarios).</p> <p>Calidad de las respuestas: si no se gestiona adecuadamente el material de entrenamiento y el proceso de mejora continua, los xatbots pueden proporcionar respuestas inadecuadas, afectando a la confianza en la información proporcionada y la percepción de la eficacia del sistema.</p>

CAMPO	RESPUESTA
<b>13. Riesgos I</b>	<p><b>Riesgos para el ciudadano:</b></p> <p>Confianza y comprensión de la asistencia: los ciudadanos pueden perder la confianza en el sistema si perciben que las respuestas no son satisfactorias o no son adecuadas para su asistencia, especialmente en casos complejos o situaciones específicas.</p> <p>Protección de datos y gestión de la información personal: existe la preocupación de que los datos personales de los usuarios puedan ser vulnerables si el sistema no gestiona adecuadamente la información recogida o si se producen errores en los protocolos de seguridad.</p> <p>Comunicación y limitaciones en la asistencia: los chatbots pueden tener limitaciones en la comprensión del lenguaje natural o en la capacidad de proporcionar respuestas adecuadas para todas las situaciones, pudiendo afectar a la calidad de la asistencia y la satisfacción del usuario.</p>
<b>14. Riesgos II</b>	<ul style="list-style-type: none"> <li>▶ Amenaza a la privacidad</li> <li>▶ Difundir información errónea</li> </ul>
<b>15. Intervención humana</b>	<p>El resultado del algoritmo es consultado por especialistas del CAU (de primer o segundo nivel) y por expertos en seguimiento del algoritmo, pero no pueden modificar la respuesta dada.</p>
<b>16. Intervención humana (detalles)</b>	<p>Los especialistas del CAU (de primer o segundo nivel) reciben las cuestiones que no se han podido responder y desarrollan los flujos de conversación que permitirán dar la respuesta adecuada la próxima vez. Los expertos en seguimiento supervisan periódicamente los registros de comentarios y conversaciones y analizan las estadísticas de uso del servicio. A partir de estos datos, los expertos identifican las cuestiones que interesan a los usuarios y desarrollan nuevos flujos de conversación y amplían el corpus conversacional con nuevas palabras clave.</p>
<b>17. Auditoría</b>	
<b>18. Transparencia</b>	<ul style="list-style-type: none"> <li>▶ Datos utilizados para entrenar el algoritmo</li> <li>▶ Tipo de algoritmo utilizado</li> <li>▶ Problemática a la cual se pretende dar solución</li> <li>▶ Cómo se está implementando</li> <li>▶ Objetivos y población que interactúa con el algoritmo</li> <li>▶ Resultados obtenidos hasta ahora</li> <li>▶ Gestión de riesgos, información de contacto para consultas y del desarrollador del algoritmo</li> </ul>
<b>19. Visibilidad</b>	<p>Se informa directamente al ciudadano (comunicación activa).</p>

CAMPO	RESPUESTA
20. Enlaces	<p>La iniciativa de publicar la ficha de transparencia algorítmica de los xatbots conversacionales de la AOC en el portal de transparencia pretende ser una de las actuaciones recomendables a las que se refiere la propuesta de reglamento de IA y las directrices de la UE para una IA fiable. La principal motivación que nos ha llevado a formular esta propuesta es velar por la transparencia de los algoritmos de IA que aplicamos en nuestros servicios de Administración digital, y para que las personas estén informadas y capacitadas para utilizarlos cuando interactúen con ellos.</p> <p>Ficha de transparencia algorítmica: <a href="https://www.aoc.cat/ia-transparencia-xatbot/">https://www.aoc.cat/ia-transparencia-xatbot/</a></p>

## 11. Servicio de videoidentificación para obtener el idCAT Móvil

CAMPO	RESPUESTA
1. Nombre	Servicio de videoidentificación para obtener el idCAT Móvil
2. Descripción	<p>La problemática inicial reside en la dificultad para las personas de obtener una identidad digital (eID) en Cataluña, especialmente durante la pandemia. Con las restricciones de movilidad y la necesidad creciente de interactuar con las administraciones desde casa, la ausencia de una identificación digital generaba obstáculos y limitaciones en las transacciones y comunicaciones con las instituciones públicas.</p> <p>La solución fue el desarrollo de un sistema de IA de videoidentificación para el idCAT Móvil. Esta solución permite a las personas obtener una identidad digital con validez oficial para interactuar con las administraciones catalanas desde cualquier lugar con acceso a internet. Mediante algoritmos de reconocimiento facial y la validación de documentos oficiales, este sistema asegura una autenticación remota fiable, equiparable a la presencia física.</p> <p>La implementación consiste en un proceso sencillo para el usuario, que puede realizar una videoidentificación desde su teléfono móvil. El sistema verifica la coincidencia entre el documento oficial presentado y el videoselfi de la persona, asegurando la integridad del proceso. Esta verificación es posteriormente validada por operadores humanos especializados para asegurar la corrección y la eliminación de posibles riesgos o problemáticas.</p> <p>Hasta ahora, este sistema ha proporcionado más de 110.000 identidades digitales con éxito, resolviendo la problemática inicial de acceso a una identidad digital durante situaciones restrictivas como la pandemia. Los resultados muestran una alta eficiencia y aceptación por parte de los usuarios, ofreciendo una solución viable y segura para sus necesidades de interacción con las administraciones sin necesidad de desplazamiento físico.</p>

CAMPO	RESPUESTA
<b>3. Objetivo I</b>	<p><b>Objetivo del sistema de IA:</b></p> <p>El principal objetivo es permitir que cualquier persona mayor de 16 años pueda obtener una identidad digital para interactuar con las administraciones catalanas desde su casa, utilizando solo un teléfono móvil y un documento oficial con fotografía, como un pasaporte o un documento nacional de identidad.</p> <p><b>Función del sistema de IA:</b></p> <p>La principal función del sistema es proporcionar un servicio de videoidentificación que, mediante el uso de algoritmos de IA, permita comprobar a distancia que una persona es quien dice ser con una elevada precisión y seguridad. Este proceso tiene como resultado el alta de la persona en el registro del sistema de identificación y firma electrónica idCAT Móvil, obteniendo así una identidad digital para poder tramitar con cualquier administración catalana con garantías de seguridad y privacidad.</p> <p>El sistema está diseñado para ser accesible 24x7, con un alto nivel de disponibilidad, y ofrece beneficios en comodidad, de inclusión y mayor seguridad; todo ello, manteniendo una supervisión humana para garantizar la corrección del proceso y cumpliendo con las normativas legales de protección de datos y seguridad.</p>
<b>4. Objetivo II</b>	<ul style="list-style-type: none"> <li>▶ Automatización de tareas</li> <li>▶ Reconocimiento de rasgos faciales</li> <li>▶ Reconocimiento de imágenes</li> <li>▶ Autenticación remota de una persona (comprobar a distancia que una persona es quien dice ser)</li> </ul>
<b>5. Adm. Responsable</b>	Administración Abierta de Cataluña (AOC)
<b>6. Entidad desarrolladora</b>	DELOITTE ADVISORY, S.L., que aporta la tecnología Deloitte OBA, integrada con la solución de VERIDAS
<b>7. Estatus</b>	Implementado: el sistema está implementado y en uso.
<b>8. Fecha de inicio</b>	Mayo de 2020
<b>9. Fecha de finalización</b>	
<b>10. Colectivo</b>	Ciudadanía mayor de 16 años

CAMPO	RESPUESTA
11. Beneficios I	<p><b>Beneficios para la Administración:</b></p> <ul style="list-style-type: none"> <li>▶ <b>Eficiencia:</b> reducción significativa del tiempo y de los costes administrativos asociados a la verificación de identidad y la gestión presencial.</li> <li>▶ <b>Seguridad:</b> aumento sustancial de la seguridad en el proceso de autenticación, con una supervisión humana que reduce riesgos de fraude o errores.</li> <li>▶ <b>Cumplimiento normativo:</b> alineación con las normativas de protección de datos y certificaciones legales para la emisión de certificados cualificados, asegurando la fiabilidad y legalidad de los documentos emitidos. Cabe decir que no existe normativa específica aplicable al idCAT Móvil, y la AOC ha decidido seguir voluntariamente el modelo de los certificados.</li> </ul> <p><b>Beneficios para el ciudadano:</b></p> <ul style="list-style-type: none"> <li>▶ <b>Comodidad:</b> acceso fácil y cómodo a una identidad digital sin necesidad de desplazamientos, con la posibilidad de completar el proceso desde casa con un teléfono móvil.</li> <li>▶ <b>Inclusión:</b> accesibilidad mejorada, con un proceso sencillo y orientado a todo tipo de personas independientemente de sus habilidades digitales.</li> <li>▶ <b>Seguridad y confianza:</b> garantía de un proceso de autenticación preciso y fiable, con un alto nivel de seguridad en la verificación de la identidad.</li> </ul>
12. Beneficios II	<ul style="list-style-type: none"> <li>▶ Servicios personalizados (la oferta de los servicios se ajusta mejor a las necesidades de la población) [a]</li> <li>▶ Servicios públicos (centrados en el ciudadano: mayor accesibilidad del ciudadano a los servicios públicos) [a]</li> <li>▶ Servicios públicos más sensibles, eficientes y rentables [a]</li> <li>▶ Nuevos servicios o canales [a]</li> <li>▶ Reducción de costes [b]</li> <li>▶ Capacidad de respuesta de la operación gubernamental [b]</li> <li>▶ Mejora de la gestión de los recursos públicos [b]</li> <li>▶ Aumento de la calidad de procesos y sistemas [b]</li> <li>▶ Permitir mayor justicia, igualdad y honestidad [b]</li> </ul>
13. Riesgos I	<p><b>Riesgos para la Administración:</b></p> <ul style="list-style-type: none"> <li>▶ <b>Protección de datos sensibles:</b> la gestión de datos biométricos y personales durante la videoidentificación puede exponer a riesgos de seguridad, especialmente si se producen interrupciones en el ciframiento o si los datos son susceptibles de ser comprometidos durante el proceso.</li> <li>▶ <b>Supervisión y control humano:</b> pese a la supervisión de los operadores, existe la posibilidad de errores humanos en la validación y verificación de los procesos, que podrían resultar en identificaciones incorrectas o en la no detección de intentos de fraude.</li> <li>▶ <b>Suplantación de identidad:</b> pese a los controles, existe el riesgo de que algunas personas puedan utilizar métodos para engañar al sistema y suplantar la identidad de otros individuos.</li> <li>▶ <b>Dependencia tecnológica:</b> la alta dependencia de este sistema podría crear vulnerabilidades en caso de fallos del sistema o interrupciones tecnológicas que afecten a la continuidad del servicio de identificación.</li> </ul>

CAMPO	RESPUESTA
13. Riesgos I	<p><b>Riesgos para el ciudadano:</b></p> <ul style="list-style-type: none"> <li>▶ <b>Privacidad y seguridad de datos:</b> el uso de datos biométricos en el proceso puede generar inquietudes sobre la privacidad y la seguridad de los usuarios, especialmente si estos datos son vulnerables a ataques cibernéticos o mal uso por terceros.</li> <li>▶ <b>Posibles sesgos en el reconocimiento facial:</b> pese a la baja tasa de rechazo, aún existe una posibilidad de sesgos en el reconocimiento facial, que podrían conducir a errores en la identificación o a casos de falsos positivos o negativos, afectando a ciertos grupos de personas de forma desigual.</li> <li>▶ <b>Dependencia de la conectividad y tecnología:</b> el proceso de videoidentificación depende de la conexión a internet y de la tecnología del dispositivo, lo que podría limitar el acceso a aquellas personas con conexiones deficientes o carencias tecnológicas.</li> </ul>
14. Riesgos II	<ul style="list-style-type: none"> <li>▶ Amenaza a la privacidad</li> <li>▶ Otros tipos de discriminación</li> </ul>
15. Intervención humana	<p>El resultado del algoritmo es supervisado por una persona que tiene la decisión final.</p>
16. Intervención humana (detalles)	<p>En concreto, las acciones que realizan los operadores que supervisan las transacciones son:</p> <ol style="list-style-type: none"> <li>a. Visualizar el vídeo y comprobar que la persona interesada ha realizado todo el proceso sin coacción y que lo ha realizado según los requerimientos establecidos (no lleva la cabeza cubierta, no lleva gafas de sol, se le ve la cara, etc.), además de revisar que se vean las marcas de agua y hologramas de autenticidad del documento mostrado.</li> <li>b. Revisar el resultado de los parámetros de evaluación de la documentación que el sistema ha evaluado de forma automática: porcentaje de coincidencia entre la foto del documento identificador y la foto selfi, fecha de caducidad del documento identificador, fecha de nacimiento, número del documento, y que sea legible y contenga las marcas de agua.</li> </ol>
17. Auditoría	<ol style="list-style-type: none"> <li>1) Para identificar los riesgos asociados a la disponibilidad y seguridad del sistema, la AOC realizó, el 7/7/2020, un análisis de seguridad del servicio de videoidentificación siguiendo las guías de la Agencia Catalana de Ciberseguridad y del Esquema Nacional de Seguridad (ENS).</li> <li>2) Para identificar los riesgos del sistema desde el punto de vista de los derechos fundamentales, los principios y valores subyacentes, la Generalitat de Catalunya realizó, el 17/12/2020, una evaluación de impacto sobre la protección de datos (APD) y una evaluación sobre la implementación del proceso de videoidentificación para obtener el idCAT Móvil.</li> <li>3) Para garantizar la imparcialidad de los algoritmos aplicados se realizan pruebas de calidad periódicas que permiten detectar imprecisiones y evitar sesgos. • Prueba de proveedores de reconocimiento facial del National Institute of Standards and Technology (NST) (PDF último informe, 16/6/2023)</li> </ol>

CAMPO	RESPUESTA
18. Transparencia	<ul style="list-style-type: none"> <li>▶ Datos utilizados para entrenar el algoritmo</li> <li>▶ Tipo de algoritmo utilizado</li> <li>▶ Problemática a la cual se pretende dar solución</li> <li>▶ Cómo se está implementando</li> <li>▶ Objetivos y población que interactúa con el algoritmo</li> <li>▶ Resultados obtenidos hasta ahora</li> </ul>
19. Visibilidad	Se informa directamente al ciudadano (comunicación activa).
20. Enlaces	<p>Ficha de innovación «Identificación remota de los ciudadanos mediante videoidentificación»: <a href="https://www.aoc.cat/projecte-innovacio/identificacio-remota-dels-ciutadans-mitjançant-videoidentificacio/">https://www.aoc.cat/projecte-innovacio/identificacio-remota-dels-ciutadans-mitjançant-videoidentificacio/</a></p> <p>La ficha de transparencia algorítmica se publicará en breve en el portal de transparencia de la AOC. [<a href="https://www.seu-e.cat/ca/web/consorciaoc/govern-obert-i-transparencia/accio-de-govern-i-normativa/normativa-plans-i-programes/transparencia-algorismica-234">https://www.seu-e.cat/ca/web/consorciaoc/govern-obert-i-transparencia/accio-de-govern-i-normativa/normativa-plans-i-programes/transparencia-algorismica-234</a>]</p>

## 12. Servicio de automatización de los informes de pobreza energética

CAMPO	RESPUESTA
1. Nombre	Servicio de automatización de los informes de pobreza energética
2. Descripción	<p>El servicio de automatización de los informes de pobreza energética es una solución desarrollada para automatizar la generación de los informes sociales de vulnerabilidad energética. Estos informes son esenciales para garantizar el acceso a los suministros básicos a personas y unidades familiares en riesgo de exclusión residencial, según la Ley 24/2015 de medidas urgentes para afrontar la emergencia en el ámbito de la vivienda y la pobreza energética (artículo 6).</p> <p>Por ese motivo, mensualmente, las diferentes áreas básicas de servicios sociales (ABSS) deben redactar los citados informes para acreditar la situación de vulnerabilidad económica ante los cortes que puedan acordar las compañías suministradoras de electricidad, gas y agua potable (más de 50 compañías diferentes en todo el territorio) por falta de pago.</p> <p>La principal problemática detectada era la complejidad y la dispersión de los datos necesarios para la elaboración de los informes de pobreza energética. Con la existencia de múltiples fuentes de datos y formatos heterogéneos, se trataba de un proceso manual propenso a errores e ineficiente en la verificación de la vulnerabilidad energética de las personas y unidades familiares en riesgo de exclusión residencial.</p>

## CAMPO

## RESPUESTA

La solución adoptada fue la implementación de un sistema de IA que utiliza algoritmos de procesamiento de datos y tecnologías como RPA y BPM. Este sistema se ha desarrollado para automatizar la recopilación y verificación de datos, la generación de los informes de pobreza energética y la comunicación del resultado a las empresas suministradoras de energía. Su implementación implica la integración con múltiples sistemas, como los servicios de interoperabilidad Via Oberta o el sistema de información de servicios sociales Hèstia.

La implementación concreta consiste en la creación de una plataforma en la nube proporcionada por una empresa privada (Consultors BPM), la cual está hospedada en Microsoft Azure dentro de la Comunidad Europea. Esta plataforma permite cargar los ficheros de los proveedores energéticos, obtener los datos socioeconómicos de los titulares y otros pasos necesarios para calcular los coeficientes y generar automáticamente los informes de vulnerabilidad. Además, se realiza la comunicación de los informes a las empresas suministradoras vía e-Notum.

Hasta ahora, los resultados muestran un porcentaje de éxito del 90 %, aunque se prevé que con cambios legales y organizativos esta cifra pueda llegar casi al 100 %.

## 2. Descripción

La implementación se ha puesto a prueba en seis municipios diferentes: El Prat de Llobregat, Ampostà, Granollers, Sant Boi de Llobregat, Palau Solità i Plegamans y Santa Coloma de Gramenet.

Los principales resultados obtenidos hasta ahora son:

1. Importante ahorro de tiempo y tareas administrativas: se ha mejorado la velocidad y eficiencia en la generación de los informes de pobreza energética, permitiendo ofrecer un mejor servicio a la ciudadanía en situación de vulnerabilidad.
2. Mayor interoperabilidad y mejora en la gestión de datos: se ha conseguido una mejora en la interconexión entre empresas, municipios y otras administraciones, facilitando una visión más precisa de la situación de vulnerabilidad energética. También una importante reducción de cargas administrativas de los ciudadanos gracias a la interoperabilidad.
3. Apoyo a los municipios con recursos limitados: el sistema ofrece apoyo a los municipios que no disponen de los recursos necesarios para grandes proyectos de transformación digital, permitiéndoles acceder a este servicio a través de un servicio en la nube sin necesidad de instalación local.

CAMPO	RESPUESTA
<p><b>3. Objetivo I</b></p>	<p>El sistema de IA tiene como principal objetivo automatizar y estandarizar la generación de los informes de pobreza energética para mejorar la gestión, la verificación de datos y la toma de decisiones en situaciones de vulnerabilidad energética; todo ello, asegurando el cumplimiento de la normativa de protección de datos y garantizando una supervisión humana para asegurar la precisión de los resultados.</p> <p>En cuanto a la función, este servicio permite:</p> <ol style="list-style-type: none"> <li>1. Estandarizar y automatizar la generación de los informes de pobreza energética.</li> <li>2. Mejorar la gestión de datos y la interoperabilidad entre múltiples fuentes y sistemas.</li> <li>3. Optimizar los recursos de los municipios para la verificación de situaciones de vulnerabilidad energética.</li> <li>4. Asegurar el cumplimiento de normativas de protección de datos y la regulación vigente.</li> </ol> <p>La supervisión humana es esencial para validar los resultados del sistema antes de enviarlos a las empresas suministradoras, asegurando la precisión y evitando posibles errores en casos específicos que puedan escapar a la lógica del sistema.</p> <p>Asimismo, el sistema está diseñado e implementado siguiendo los requisitos de la normativa de protección de datos, especialmente el Reglamento General de Protección de Datos (RGPD) y otras regulaciones locales. Cumplir con estos estándares es crucial, ya que se trabaja con datos personales, y debe garantizarse la transparencia, la limitación en la recopilación de datos y el uso correcto del consentimiento de las personas afectadas.</p>
<p><b>4. Objetivo II</b></p>	<ul style="list-style-type: none"> <li>▶ Automatización de tareas</li> <li>▶ Estandarización de procesos, simplificación y reducción de cargas administrativas</li> </ul>
<p><b>5. Adm. Responsable</b></p>	<p>Administración Abierta de Cataluña (AOC)</p>
<p><b>6. Entidad desarrolladora</b></p>	<p>Consultors BPM</p>
<p><b>7. Estatus</b></p>	<p>Piloto: el sistema está siendo probado.</p>
<p><b>8. Fecha de inicio</b></p>	<p>27 de enero de 2023</p>
<p><b>9. Fecha de finalización</b></p>	
<p><b>10. Colectivo</b></p>	<p>Personal de las áreas básicas de servicios sociales (ABSS)</p>

CAMPO	RESPUESTA
<p><b>11. Beneficios I</b></p>	<p><b>Beneficios para la Administración:</b></p> <p>Ahorro de tiempo y recursos: reducción significativa del tiempo y las tareas administrativas necesarias para generar los informes de pobreza energética. Eso permite una mejor gestión de los recursos humanos y económicos de la Administración, optimizando su rendimiento.</p> <p>Estandarización y mejora en la gestión de datos: a través de la integración de múltiples fuentes de datos y su automatización, se consigue una visión más precisa y estandarizada de la situación de vulnerabilidad energética. Eso mejora la toma de decisiones y facilita la comunicación entre empresas, municipios y otras administraciones.</p> <p>Apoyo a los municipios con recursos limitados: el servicio en la nube ofrece la posibilidad de acceso a una tecnología avanzada sin la necesidad de grandes inversiones en infraestructura o recursos internos.</p> <p><b>Beneficios para el ciudadano:</b></p> <p>Mejor servicio y agilidad en la atención: la automatización y estandarización del proceso permite una respuesta más rápida y precisa a los casos de vulnerabilidad energética, mejorando la calidad de los servicios sociales y la atención a las personas en situación de riesgo.</p> <p>Asegurar el acceso a los suministros básicos: a través de este sistema, se agiliza la determinación de la situación de vulnerabilidad, asegurando que las personas y unidades familiares en riesgo reciban la atención y los suministros básicos necesarios.</p> <p>Reducción de las cargas administrativas y mejora de la experiencia del usuario: el proceso automatizado evita trámites innecesarios y simplifica la interacción con la Administración, reduciendo la burocracia y mejorando la experiencia de los ciudadanos en situaciones de vulnerabilidad.</p>
<p><b>12. Beneficios II</b></p>	<ul style="list-style-type: none"> <li>▶ Servicios públicos más sensibles, eficientes y rentables [a]</li> <li>▶ Reducción de costes [b]</li> <li>▶ Capacidad de respuesta de la operación gubernamental [b]</li> <li>▶ Mejora de la gestión de los recursos públicos [b]</li> <li>▶ Aumento de la calidad de procesos y sistemas [b]</li> <li>▶ Mejor colaboración y mejor comunicación [b]</li> <li>▶ Reducir o eliminar el riesgo de corrupción y abuso de la ley por parte de los funcionarios públicos [b]</li> <li>▶ Permitir mayor justicia, igualdad y honestidad [b]</li> </ul>

CAMPO	RESPUESTA
<p data-bbox="188 1061 355 1093"><b>13. Riesgos I</b></p>	<p data-bbox="448 371 1359 465">La implementación de este sistema de IA, pese a sus ventajas, también puede presentar algunos riesgos, tanto para la Administración como para los ciudadanos.</p> <p data-bbox="448 497 880 528"><b>Riesgos para la Administración:</b></p> <ul data-bbox="448 537 1359 864" style="list-style-type: none"> <li data-bbox="448 537 1359 667">▶ Dependencia de la tecnología: si algo falla en la plataforma o en el proceso automatizado, podría provocar retrasos o errores en la generación de los informes, poniendo presión adicional sobre los recursos humanos para solucionarlos.</li> <li data-bbox="448 676 1359 864">▶ Vulnerabilidad y seguridad de datos: la integración de múltiples fuentes de datos y el uso de servicios en la nube pueden generar vulnerabilidades de seguridad. Si el sistema no está suficientemente protegido, podría ser susceptible de sufrir ataques cibernéticos o pérdida de datos, comprometiendo la privacidad y la seguridad de la información de los ciudadanos.</li> </ul> <p data-bbox="448 891 823 922"><b>Riesgos para el ciudadano:</b></p> <ul data-bbox="448 931 1359 1151" style="list-style-type: none"> <li data-bbox="448 931 1359 1151">▶ Posibles errores en la evaluación de la vulnerabilidad y pérdida de confianza: pese a la supervisión humana, existe la posibilidad de que el sistema cometa errores en casos específicos. Eso podría resultar en informes inexactos sobre la vulnerabilidad energética de las personas o unidades familiares, afectando su capacidad de acceder a los suministros básicos y provocando una pérdida de confianza en la Administración.</li> </ul> <p data-bbox="448 1169 1359 1608">La pérdida de confianza de los ciudadanos en la Administración es un factor crítico que puede surgir si se producen errores frecuentes en la evaluación de la vulnerabilidad. Esos errores pueden generar desconfianza en la fiabilidad de los informes generados, afectar la percepción de transparencia y justicia en la toma de decisiones y tener un impacto emocional negativo en las personas afectadas. Si se deniega el acceso a los suministros básicos debido a informes incorrectos, pueden provocarse situaciones de desconfianza y desesperación en aquellos que ya se encuentran en situaciones de vulnerabilidad. Por eso, es crucial gestionar este riesgo mediante una supervisión minuciosa, procesos transparentes y una comunicación efectiva para no perder la confianza de los ciudadanos en los servicios que ofrece la Administración en situaciones de vulnerabilidad energética.</p> <ul data-bbox="448 1626 1359 1787" style="list-style-type: none"> <li data-bbox="448 1626 1359 1787">▶ Problemas de privacidad y consentimiento: trabajar con datos personales implica el riesgo de violación de la privacidad, si no se gestionan adecuadamente. Debe garantizarse que se obtenga y se utilice el consentimiento de las personas afectadas de forma transparente y conforme a la normativa de protección de datos.</li> </ul>
<p data-bbox="188 1809 363 1841"><b>14. Riesgos II</b></p>	<p data-bbox="448 1809 801 1841">Amenaza a la privacidad.</p>
<p data-bbox="188 1863 402 1926"><b>15. Intervención humana</b></p>	<p data-bbox="448 1863 1359 1926">El resultado del algoritmo es supervisado por una persona que tiene la decisión final.</p>
<p data-bbox="188 1953 437 2016"><b>16. Intervención humana (detalles)</b></p>	

CAMPO	RESPUESTA
<b>17. Auditoría</b>	<ul style="list-style-type: none"> <li>▶ Problemática a la cual se pretende dar solución</li> </ul>
<b>18. Transparencia</b>	<ul style="list-style-type: none"> <li>▶ Cómo se está implementando</li> <li>▶ Resultados obtenidos hasta ahora</li> </ul>
<b>19. Visibilitat</b>	<p>La información es pública pero poco accesible (depende de la proactividad del ciudadano para informarse).</p>
<b>20. Enlaces</b>	<ul style="list-style-type: none"> <li>▶ Post «Piloto para la automatización del informe de pobreza energética para evitar cortes de suministros» (27/1/2023, blog AOC): <a href="https://www.aoc.cat/blog/2023/pilot-pobresa-energetica/">https://www.aoc.cat/blog/2023/pilot-pobresa-energetica/</a></li> <li>▶ Artículo «Automating the energy poverty report in Catalunya» (10/10/2023, Plataforma Joinup de la Comisión Europea): <a href="https://joinup.ec.europa.eu/collection/public-sector-tech-watch/automating-energy-poverty-report-catalunya">https://joinup.ec.europa.eu/collection/public-sector-tech-watch/automating-energy-poverty-report-catalunya</a></li> <li>▶ Presentación «Evaluación automatizada de la pobreza energética» (18/10/2023, Semic 2023: Interoperable AI in the Age of AI): <a href="https://www.linkedin.com/posts/miquelestape_20231018-semic-2023-trustworthy-ai-for-activity-7122219700981256194-f9es/?utm_source=share&amp;utm_medium=member_desktop">https://www.linkedin.com/posts/miquelestape_20231018-semic-2023-trustworthy-ai-for-activity-7122219700981256194-f9es/?utm_source=share&amp;utm_medium=member_desktop</a></li> <li>▶ La ficha de transparencia algorítmica se publicará en breve en el portal de transparencia de la AOC. [<a href="https://www.seu-e.cat/ca/web/consorcioaoc/govern-obert-i-transparencia/accio-de-govern-i-normativa/normativa-plans-i-programes/transparencia-algorismica-234">https://www.seu-e.cat/ca/web/consorcioaoc/govern-obert-i-transparencia/accio-de-govern-i-normativa/normativa-plans-i-programes/transparencia-algorismica-234</a>]</li> </ul>

## 5. VALORACIONES FINALES

En este segundo documento del proyecto, **el objetivo prioritario ha sido la identificación de sistemas de inteligencia artificial y procesos automatizados en el ámbito social que estuviesen siendo utilizados, o en fase piloto, por parte de las administraciones públicas catalanas.**

Para conseguir este objetivo fue necesario, por un lado, definir con claridad qué información quería recogerse. Es decir, los campos del radar. Por otro lado, fue necesario llevar a cabo un minucioso trabajo de campo identificando tanto personas como instituciones y departamentos que pudiesen tener conocimiento o estar utilizando estos sistemas. Como resultado de ese esfuerzo, **se consiguieron mapear hasta un total de 12 algoritmos**, que muy probablemente son (si no todos) la mayor parte de los sistemas de IA utilizados actualmente en Cataluña o que están en fase piloto.

En este punto **es imprescindible poner en valor la colaboración de los profesionales de la Administración pública.** Los técnicos de la Generalitat de Cataluña, del Ayuntamiento de Barcelona, del Instituto Municipal de Servicios Sociales y de la Administración Abierta de Cataluña respondieron el cuestionario proporcionando un notable grado de detalle sobre los sistemas de IA, lo que ha permitido construir el radar.

Además, en las conversaciones con los técnicos quedaron patentes las particularidades del ámbito de los derechos sociales. La obligación de diseñar procesos con garantías a la hora de proveer servicios al ciudadano limita considerablemente el abanico de técnicas de inteligencia artificial que pueden aplicarse. Por ejemplo, si se utilizase un sistema de IA para decidir si conceder o no una prestación a una persona, ese sistema no debería tratarse de un algoritmo de caja negra (como una red neuronal o un *random forest*), dado que, al menos hoy día, la explicabilidad de esos modelos es baja y, por lo tanto, la capacidad de explicar la decisión del modelo también lo es. Para una aplicación de este tipo, sería preferible recurrir a automatizaciones (estrictamente basadas en la regulación) o algoritmos explicables (como las regresiones, habitualmente utilizadas en economía o ciencia política).

Los técnicos transmitieron en todo momento que la obligación de dar respuestas claras a los ciudadanos sobre el porqué de un resultado con-

dicionaba el hecho de adoptar o no este tipo de sistemas. Eso queda claramente reflejado en el radar, donde buena parte de los algoritmos son de «consumo propio», es decir, están dirigidos a automatizar procedimientos o a ayudar a los técnicos de la Administración, pero no a tomar decisiones que afecten directamente al ciudadano. Es razonable esperar que los tipos de soluciones que adopten las administraciones públicas en el ámbito social difieran en muchos casos de las que puedan tomarse en otras áreas.

A modo de reflexión final, muy probablemente estamos ante un cambio de paradigma. Y en todo proceso de transformación social profundo, es habitual que parte de la sociedad se divida entre el temor por la incertidumbre futura y un optimismo excesivo y acrítico ante los cambios venideros. No es sencillo encontrar el punto medio que contribuya a un desarrollo efectivo y justo. Es decir, una perspectiva que opte por analizar caso por caso de forma rigurosa y sin apriorismos. **El valor fundamental de este trabajo es que ha ayudado a obtener aprendizajes importantes a la hora de comprender cómo se están aplicando actualmente los sistemas de inteligencia artificial en el ámbito social.**

En particular, ha permitido extraer una serie de conclusiones clave. En primer lugar, **el número total de sistemas identificados es reducido, un total de 12**. Diversos factores pueden ayudar a entender por qué en este ámbito existe un menor número de sistemas que en otros, como el de la salud. Probablemente, la disponibilidad de recursos sea uno de ellos. No obstante, la relación entre la Administración pública y el ciudadano en el ámbito social sigue una lógica de garantías, por lo que el abanico de opciones donde implementar aplicaciones de inteligencia artificial se ve reducido.

En segundo lugar, y derivado de lo anterior, **buena parte de estos sistemas son de «consumo interno»**, es decir, quien interactúa con ellos es el personal de la Administración, no el ciudadano. De esta forma, estos sistemas están **más centrados en producir mejoras en la eficiencia de los procesos administrativos**.

En tercer lugar, **los sistemas identificados comportan niveles de riesgo muy reducidos o inexistentes**. Los algoritmos aplicados están dirigidos a resolver problemáticas que, dada su naturaleza, apenas comportan ningún riesgo para el ciudadano.

En cuarto lugar, **la colaboración de los responsables de la Administración pública ha sido positiva**. Eso ha permitido que la mayor parte de la información recogida provenga directamente de personas que han participado en el diseño, gestión o aplicación de estos sistemas.

Finalmente, **no puede perderse de vista que la inteligencia artificial es una herramienta más, y solo es la más óptima, eficiente o efectiva para algunas problemáticas determinadas**. No es un fin en sí mismo. Desde la Administración pública debe valorarse si, para cada problema, existe un sistema alternativo a la IA que sea mejor. Esta concepción fue percibida con claridad en las conversaciones mantenidas con las personas técnicas.





Taula d'entitats  
del Tercer Sector Social  
de Catalunya

m4Social

Con el apoyo de:



Generalitat de Catalunya  
Departament  
de Drets Socials



**AMB**

Àrea Metropolitana  
de Barcelona