

Uso y aplicación de la inteligencia artificial en políticas públicas de empleo e inclusión social: una revisión exploratoria¹

Beatriz Vallina Acha

Universitat Politècnica de València
beatriz.vallina.acha@gmail.com

Barbara Branchini

Fresno Servicios Sociales, S.L.
barbara.branchini@fresnoconsulting.es

Leticia Henar Lomeña

Fresno Servicios Sociales, S.L.
leticia.henar@fresnoconsulting.es

Adimen artifiziala sektore publikoan aplikatzeko gero eta interes handiagoa dagoen arren, oraindik ez du ebidentzia sistematizaturik enplegu- eta gizartratze-politiketan erabiltzeak. Azterketa honek politikak ezartzeko aplikazioan gaur egun duen eta sortzen ari den egoera mapeatzen du; esploratzeko berrikuspen bibliografiko baten eta 25 azterlanen eta literatura analitikoaren azterketa tematikoaren bitartez. Emaitez erakusten dutenez, adimen artifiziala prozesu hauetan erabiltzen da: automatizazio administratiboa, erabiltzaileen sailkapena, kalteberatasun-egoerak alde aurretik hautematea eta herritarren parte-hartzeari laguntzea. Erabilitako teknologien artean prozesuen automatizazio robotikoa, ikaskuntza automatikoa, lengoia naturalaren prozesamendua eta erabaki automatizatuaren sistemak daude. Berrikusatiko kasuek agerian uzten dituzte onura operatibo nabariak eta hobekuntzak irisgarritasunean eta pertsonalizazioan, baina baita tentsio etiko eta operatiboak ere. Sistema algoritmiko konplexuagoetarantz eboluzionatzeko, beraz, gardentasuna, ekitatea eta kontrol instituzionala bermatuko duten gobernantza-esparruak behar dira.

Gako-hitzak:

Adimen artifiziala, politika publikoak, gizarte-zerbitzuak, enplegu-zerbitzuak, gobernantza algoritmikoa.

A pesar del creciente interés por aplicar la inteligencia artificial en el sector público, su uso en las políticas de empleo e inclusión social aún carece de evidencia sistematizada. Este estudio mapea el estado actual y emergente de su aplicación en la implementación de políticas, mediante una revisión bibliográfica exploratoria y el análisis temático de 25 estudios y literatura analítica. Los resultados muestran que la inteligencia artificial se emplea en procesos de automatización administrativa, clasificación de personas usuarias, detección anticipada de situaciones de vulnerabilidad y apoyo a la participación ciudadana. Las tecnologías utilizadas incluyen automatización robótica de procesos, aprendizaje automático, procesamiento de lenguaje natural y sistemas de decisión automatizada. Los casos revisados evidencian beneficios operativos tangibles y mejoras en la accesibilidad y personalización, pero también tensiones éticas y operativas. La evolución hacia sistemas algorítmicos más complejos exige, por tanto, marcos de gobernanza que garanticen transparencia, equidad y control institucional.

Palabras clave:

Inteligencia artificial, políticas públicas, servicios sociales, servicios de empleo, gobernanza algorítmica.

¹ Esta investigación forma parte del estudio de *Exploración, análisis y prospección en la aplicación de la IA en los procesos de evaluación de las políticas públicas*, elaborado por Fresno Consulting previo encargo de Besaldi-Órgano de Evaluación de las Políticas de Empleo e Inclusión. El informe completo y un recopilatorio de 22 casos de uso pueden consultarse en: <<https://www.euskadi.eus/besaldi/documentos/web01-a2besald/es/>>.

1. Introducción

En los últimos años, varias administraciones públicas del entorno europeo han empezado a utilizar herramientas de inteligencia artificial (IA), a la vez que su aplicación se está ampliando y creciendo de forma exponencial (Tangi *et al.*, 2022). Por ello, los sistemas de IA que pretenden servir al público o al bien social constituyen un asunto de relieve para la investigación, y un objetivo explícito de muchas estrategias y propuestas regulatorias nacionales e internacionales (Züger y Asghari, 2023).

La OCDE define la IA como sistemas basados en máquinas que infieren a partir de datos, generando predicciones o contenidos capaces de influir en entornos físicos o virtuales, con niveles variables de autonomía y adaptabilidad (Recommendation of the Council on Artificial Intelligence, 2024). La Estrategia Española de Inteligencia Artificial destaca su capacidad para resolver problemas complejos mediante técnicas avanzadas que superan la necesidad de instrucciones predefinidas (Ministerio para la Transformación Digital y de la Función Pública, 2024). El estándar ISO/IEC 22989 concibe la IA como mecanismos que emulan aspectos de la inteligencia humana para generar resultados orientados a objetivos definidos por humanos (ISO e IEC, 2022). La Comisión Europea (2018) también enfatiza la capacidad de análisis del entorno y acción autónoma orientada a metas humanas

Funcionalmente, la IA puede entenderse como un software que toma decisiones de manera similar a la inteligencia humana. A diferencia de la programación tradicional, donde el desarrollador especifica cada paso computacional, los sistemas de IA aprenden patrones mediante datos sin seguir reglas explícitas (Moyano-Arias *et al.*, 2024). El aprendizaje automático representa un cambio fundamental: los procesos de resolución no están determinados de antemano, sino que emergen del entrenamiento, permitiendo que los sistemas se perfeccionen sin, necesariamente, modificar su código (Goodfellow *et al.*, 2016).

Esta capacidad de aprendizaje y adaptación plantea desafíos específicos en verificación, validación y explicabilidad, reconocidos como elementos esenciales (véase el estándar ISO/IEC 22989). En modelos complejos, estos procesos requieren enfoques especializados para garantizar la confiabilidad, la transparencia y la seguridad (ISO e IEC, 2022). El mecanismo de adaptación refuerza la centralidad de los datos y conecta con el problema de la opacidad; los modelos resultan difíciles de interpretar incluso para sus diseñadores, fenómeno conocido como el problema de la "caja negra" (Carabantes, 2020), como se verá. La explicabilidad, interpretabilidad o XAI (Barredo Arrieta *et al.*, 2020) de los modelos resulta crucial en aplicaciones donde la transparencia y la rendición de cuentas son fundamentales, como en el ámbito de la acción pública.

En el sector público, la adopción de la IA es un fenómeno en expansión que ha experimentado un crecimiento sostenido, orientado especialmente a la mejora de los servicios y la eficiencia administrativa (Comisión Europea, 2024). No obstante, la evidencia empírica sobre su uso e impacto es escasa y basada esencialmente en estudios de caso y experiencias autorreportadas (OCDE, 2025). La necesidad de experiencias documentadas es aún más relevante cuando se abordan programas de inclusión social, que implican poblaciones vulnerables y efectos a largo plazo, un ámbito en el que, sin embargo, se registra una cantidad limitada de estudios y de referencias indexadas (Raya Diez *et al.*, 2021).

Esta cuestión plantea la necesidad de sistematizar las aplicaciones existentes de la IA en las políticas públicas, sus implicaciones y potencialidades emergentes. Nuestro estudio intenta responder a esta necesidad a través de una revisión bibliográfica exploratoria, cuyo objetivo principal es detectar, analizar y sistematizar aplicaciones existentes, particularmente en contextos administrativos europeos comparables al español, y enmarcados en el ciclo de políticas públicas, con especial atención a los ámbitos del empleo y la inclusión social. Estos ámbitos representan pilares esenciales del bienestar ciudadano que enfrentan transformaciones sustanciales en el actual contexto de cambio tecnológico, económico y social. El estudio de la aplicación potencial de la inteligencia artificial en estos ámbitos cobra especial relevancia ante la necesidad de desarrollar respuestas innovadoras frente a problemas sociales emergentes y de optimizar la eficacia y eficiencia de la intervención pública en un contexto de recursos limitados y necesidades crecientes, así como su aplicación equitativa y sostenible.

La revisión exploratoria se ha centrado en la fase de implementación, entendida como el "conjunto de procesos que, tras la fase de programación, tienden a la realización concreta de los objetivos de una política pública" (Subirats, 2008: 180). La implementación se centra en la puesta en práctica de acciones concretas y produce actos formales que se destinan a las personas que forman parte de los grupos-objetivo previstos (Subirats, 2008). De este modo, la exploración bibliográfica aborda tanto la aplicación de la IA a los procesos operativos de coordinación entre actores como los servicios entregados que materializan los fines de la política.

El estudio, de carácter analítico-descriptivo, se ha vertebrado en torno a la siguiente pregunta de investigación: ¿cómo se está aplicando actualmente la IA en la implementación de políticas públicas de empleo e inclusión social en el ámbito europeo y en sistemas administrativos comparables/ similares al español? La contribución específica de esta investigación radica en proporcionar un mapeo sistemático de aplicaciones de la IA en la implementación de políticas públicas de empleo e inclusión social, generando así una base de

conocimiento estructurada para informar tanto la gestión pública como el desarrollo de una agenda de investigación futura.

2. Materiales y métodos

Este estudio es una revisión exploratoria (*scoping review*), pues debido a su pregunta de investigación, esta metodología se consideró idónea para abordar el estado actual de la aplicación de la IA en la implementación de políticas públicas. Además, el carácter incipiente y evolutivo de la integración de tecnologías disruptivas en la administración pública presenta experiencias pioneras que coexisten con extensos territorios de aplicación potencial. La revisión exploratoria constituye un tipo de revisión bibliográfica orientado a explorar la evidencia disponible sobre un asunto, con el propósito fundamental de mapear conceptos, detectar recursos e investigaciones realizadas en diversos contextos, tanto académicos como no académicos, para detectar oportunidades y vías de análisis y aplicación. Para su realización, se siguió el marco metodológico originalmente propuesto por Arksey y O'Malley (Levac *et al.*, 2010), con las clarificaciones y mejoras introducidas por Levac (2010).

2.1. Criterios de inclusión y exclusión

El foco temático comprende el estado actual del uso de la IA en la implementación de políticas sociales, con énfasis particular en intervenciones relacionadas con el empleo y la inclusión social. Si bien el objetivo primordial ha sido mapear aplicaciones existentes o en curso documentadas en la literatura, se han detectado ciertas tendencias señaladas como emergentes en los propios estudios analizados y direcciones futuras para su aplicación, que deben entenderse como resultado derivado del análisis de la literatura existente, no como producto de un ejercicio prospectivo formal. La selección documental se ha basado en parámetros claramente delimitados. Se incluyeron publicaciones académicas (artículos y monografías), estudios revisados por pares (incluyendo tesis doctorales) y literatura gris gubernamental y del tercer sector siempre que reflejaran casos de uso concretos y transferibles, así como documentos institucionales, planes estratégicos y diagnósticos. El marco temporal se definió desde 2019 hasta la actualidad, decisión fundamentada en el avance técnico, la adopción institucional y ciertos hitos relevantes (por ejemplo, los documentos de consenso ALTAI sobre ética de la IA y desarrollos legislativos europeos y nacionales). Se consideraron estudios con diversas aproximaciones metodológicas (cuantitativos, cualitativos, mixtos y, con ciertas reservas, teóricos), procedentes de bases de datos académicas y generalistas (Google Académico, Google, SIIS-Servicio de Información e Investigación Social, WOS y SCOPUS), en español e inglés. En contraposición, quedaron excluidos los artículos periodísticos, entradas de blog, publicaciones breves,

informes relativos a políticas educativas, de salud o infraestructuras, así como aquellos referentes a ámbitos de negocio (marketing, finanzas) o anteriores a 2019.

Dentro del ámbito temático, se revisaron documentos que abordan explícitamente aplicaciones de IA en la implementación de políticas sociales, su efecto y casos de uso en programas sociales. Por tanto, la búsqueda se centró en términos como *artificial intelligence, machine learning, social policy, welfare policy, public administration, government program, social work*, social service*, minimum income, guaranteed income, basic income, income support, employment, employability, labour inclusion, implementation, case study, application o evaluation*. La búsqueda sistemática se realizó en los motores y fuentes detallados previamente mediante algoritmos booleanos replicables y estructurados², adaptados a fuentes tanto académicas como generalistas. Esta duplicidad de fuentes resulta oportuna para cubrir también literatura institucional y casos de uso probados y documentados en la práctica.

2.2. Procedimiento de selección

El procedimiento para la selección final de publicaciones y extracción de información comprendió varias fases secuenciales:

1. Exploración preliminar de resultados mediante *skimming* (lectura diagonal) de títulos y resúmenes (*abstracts*), que permite una valoración rápida de la potencial relevancia; no constituye criterio suficiente de inclusión.
2. Aplicación de criterios de inclusión y exclusión: tipo de publicación, aproximación metodológica, temporalidad, idiomas y alcance geográfico, incluyendo literatura gris.
3. Priorización basada en el criterio de relevancia, operacionalizado en una escala categórica de 0 (relevancia mínima/no relevante) a 4 (relevancia alta) sobre los *abstract*.
4. Revisión detenida de la literatura sobre artículos muy relevantes (puntuación 3-4), respondiendo a principios de relevancia contextual, casos de uso, transferibilidad y actualidad.

El criterio de relevancia permitió detectar estudios con mayor potencial analítico para su revisión exhaustiva, priorizando casos de implementación práctica en el contexto europeo. Su operacionalización se estableció mediante una escala categórica de cuatro niveles.

² Una descripción detallada y rigurosa de los algoritmos de búsqueda puede consultarse en el anexo I del informe completo (Fresno Consulting, 2025b: 54-57).

Cuadro 1. Criterio de selección (relevancia)		
Puntuación	Nivel de relevancia	Descripción
0	No relevante	Satisface criterios formales de inclusión, pero presenta contenido tangencial o excesivamente teórico sin conexión operativa con la aplicación de la IA en políticas públicas de empleo o inclusión social.
1	Poco relevante	Aborda la implementación de IA desde perspectivas generalistas, sin especificidad contextual europea o con aplicabilidad restringida en el marco del empleo y la inclusión.
2	Relevante	Incorpora elementos significativos sobre aplicación de la IA en estas políticas, con aportaciones teórico-empíricas valiosas, aunque con orientación predominantemente conceptual.
3-4	Muy relevante	Presenta pertinencia crítica sobre la implementación de la IA en políticas de empleo e inclusión, con casos específicos, contribuyendo sustancialmente a comprender aplicaciones prácticas de la IA en estos ámbitos dentro del contexto europeo.

Fuente: elaboración propia

El análisis temático de las fuentes seleccionadas se realizó mediante la herramienta de análisis cualitativo MAXQDA (versiones 2024 y 2022) y, como apoyo al procesamiento sistemático de información, se utilizaron herramientas de IA generativa (Claude Sonnet 3.7, y Gemini 2.5 Pro) para determinadas tareas de categorización y síntesis preliminar, siempre bajo supervisión humana y verificación cruzada de contenidos. Adicionalmente, se empleó Claude Sonnet 4.5 con estilos personalizados para tareas específicas de redacción, enfocadas a mejorar la claridad expositiva, estructura argumentativa y continuidad narrativa del texto.

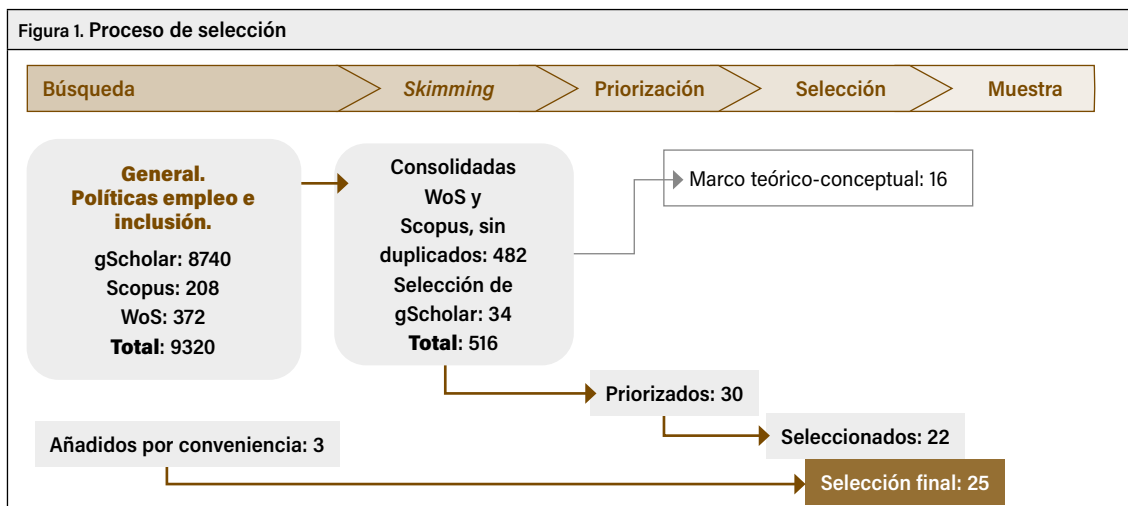
Para el apoyo con Claude Sonnet 3.7 Extended en la revisión de artículos, se implementó un proyecto en la misma plataforma, denominado ClaudeReview, para el procesamiento asistido de documentos seleccionados. Se utilizó para procesar documentos completos y extraer información estructurada según una plantilla predefinida que incluía campos esenciales como información bibliográfica, tema central, enfoque metodológico, contexto geográfico e institucional, descripción de la intervención, tipo de IA aplicada, resultados documentados, limitaciones identificadas y lecciones aprendidas. El proyecto también permitía extraer información ordenada a

partir de tablas exportadas desde MAXQDA tras codificación completamente manual (sin usar la herramienta de IA integrada), así como directamente de artículos e informes en pdf.

2.3. Artículos seleccionados

El proceso de selección de la literatura relevante para esta revisión exploratoria se desarrolló mediante un enfoque sistemático y secuencial: se combinó búsqueda estructurada en bases de datos académicas con estrategias complementarias, asegurando la exhaustividad y pertinencia de la información recogida.

Se realizaron búsquedas específicas en fuentes académicas: Google Académico (8740 resultados), Scopus (208 resultados) y WOS (372 resultados). En total, se vieron 9320 resultados. Tras el *skimming* —esto es, la lectura diagonal de título y resumen—, quedaron 516, de los cuales se priorizaron 30 para análisis detallado. El proceso de refinamiento final condujo a la selección de 22 estudios, a los que se añadieron 3 más por conveniencia, fruto de otra iteración centrada en el uso de la IA en evaluación de políticas públicas. La muestra definitiva quedó constituida, por tanto, por 25 estudios.



gScholar: Google Académico. WoS: Web of Science.
Fuente: elaboración propia

Es pertinente señalar que no todos los documentos incluían necesariamente casos de uso específicos o diferenciados, y que a lo largo de la revisión se detectaron 16 referencias de naturaleza predominantemente teórico-conceptual que aparecían citadas de manera recurrente.

2.4. Limitaciones

La metodología empleada en esta revisión exploratoria presenta un enfoque estructurado y teóricamente sólido aun presentando restricciones inherentes al tipo de estudio. La efectividad de combinar bases académicas con motores generalistas queda comprometida por los algoritmos de búsqueda, que priorizan resultados según lógicas no necesariamente académicas; esto potencialmente sesgó los resultados hacia ciertos tipos documentales. Asimismo, el desfase temporal entre publicación y aparición en resultados de búsqueda excluyó contribuciones muy recientes en un campo de rápida evolución. Un factor crítico concierne a los contextos geográficos examinados, condicionados por limitaciones lingüísticas (y el enfoque en transferibilidad al contexto español), lo que podría haber invisibilizado desarrollos relevantes en regiones no anglófonas o hispanohablantes.

La exclusión de fuentes periodísticas y blogs, justificada por el rigor empírico buscado, probablemente omitió indicadores tempranos de innovación que suelen funcionar como “radares”. Valga como ejemplo el hecho de que publicaciones significativas como el reciente informe de la Comisión Europea (Dirección General de Empleo, Asuntos Sociales e Inclusión e ICF, 2025) o los análisis de la OCDE (Brioscú *et al.*, 2024) no fueron incluidos, si bien ofrecen perspectivas complementarias sobre la evolución y tecnologías avanzadas en el sector público.

Del mismo modo, la literatura gris más reciente ofrece ejemplos significativos sobre la implementación de IA en servicios públicos de empleo (SPE) —iniciativas como EMI en Galicia (Servicio Público de Empleo Estatal, 2024) o InfoSIA en el Ayuntamiento de Madrid (Asociación ASLAN, 2025)— que no fueron detectados durante el proceso de revisión sistemática. Estos casos no forman parte del corpus de evidencia analizado, y se mencionan aquí únicamente a modo ilustrativo de lo que puede quedar fuera del alcance metodológico y también como reflejo del conocimiento tácito y experto del equipo investigador, que complementa la evidencia formal documentada.

De esta manera, en la búsqueda de la replicabilidad científica se sacrifica hasta cierto punto la exhaustividad para captar todo el conocimiento relevante; un enfoque más robusto habría incluido una fase preliminar de revisión narrativa menos estructurada, complementada con consultas a expertos para extraer conocimiento tácito y mejorar la sensibilidad semántica.

3. Resultados

Del análisis de los artículos seleccionados, emergen tensiones y complementariedades significativas en esa complejidad que implica la transformación digital, en general, y en el ámbito público, en particular; un panorama complejo caracterizado por transformaciones en los sistemas tecnológicos y en la sociedad, y tensiones que surgen al implementar la inteligencia artificial en el sector público, entre estandarización y personalización, evolución tecnológica, estructuras organizativas y marcos regulatorios, desigualdad y asimetría de recursos, opacidad y transparencia.

El compendio de casos de uso, tecnologías empleadas y principales riesgos identificados se presenta en el Anexo I. El análisis de las experiencias documentadas revela un ecosistema tecnológico diversificado que se nutre principalmente del aprendizaje automático (*machine learning*, ML), tanto en sus variantes *supervisada* (SML), donde se entrena al sistema para replicar codificaciones o clasificaciones humanas, como *no supervisada* (UML), empleada a menudo para identificar patrones o agrupar conceptos en grandes volúmenes de datos sin etiquetas previas. Junto con el *machine learning*, encontramos aplicaciones relevantes del procesamiento del lenguaje natural, la IA generativa y el análisis visual.

La revisión de la literatura revela cuatro categorías principales de aplicación de la IA en el ámbito de las políticas públicas de empleo e inclusión social en el contexto europeo: eficiencia administrativa, sistemas de perfilado y segmentación, sistemas de detección y respuesta temprana, y participación ciudadana.

La eficiencia administrativa mediante automatización constituye el marco predominante de implementación actual. Se observa una progresión desde funciones administrativas básicas (automatización robótica de procesos, comunicaciones automatizadas y gestión documental) hacia sistemas más sofisticados que apoyan la toma de decisiones, manteniendo la supervisión humana como elemento central. Los casos analizados en países como Suecia, Finlandia, Italia y Alemania demuestran cómo estas tecnologías están transformando la prestación de servicios públicos, con particular énfasis en la incorporación de asistentes virtuales y chatbots que facilitan el acceso ciudadano a información y servicios.

Por otro lado, el reconocimiento de patrones mediante aprendizaje automático ha permitido desarrollar sistemas de clasificación y priorización de poblaciones objetivo. Estos sistemas, implementados en países como Austria y Finlandia, utilizan algoritmos predictivos para evaluar características personales y trayectorias laborales, permitiendo una segmentación más eficiente de las personas usuarias de los servicios públicos. Sin embargo, estos desarrollos plantean importantes desafíos éticos relacionados con la transparencia algorítmica y la potencial

discriminación, que intentan mitigarse, como se verá, mediante la incorporación del paradigma *human-in-the-loop*.

Finalmente, un área de creciente desarrollo es la aplicación de algoritmos predictivos para la detección proactiva de necesidades sociales y situaciones de vulnerabilidad. Desde la protección infantil en Dinamarca hasta la predicción de la pobreza energética en los Países Bajos, estas aplicaciones buscan anticiparse y responder oportunamente a riesgos sociales. No obstante, casos como el sistema implementado en el municipio danés de Gladsaxe evidencian las controversias relacionadas con la privacidad y el tratamiento de datos sensibles, subrayando la complejidad ética inherente a estas aplicaciones tecnológicas en el ámbito social.

En cuanto al ámbito territorial de referencia, cabe señalar que muchos estudios analizados son revisiones y compilaciones de casos y presentan, por tanto, casuísticas variadas y pluriterritoriales, lo cual dificulta una cuantificación específica de casos por país. Por ello, hemos analizado con búsqueda simple en MAXQDA las referencias a países y la frecuencia en documentos, y hemos obtenido un resultado coherente en gran medida con estudios previos (Tangi *et al.*, 2022: 35).

3.1. Eficiencia administrativa y automatización

El vinculado de una u otra forma a la eficiencia y la automatización creciente de la Administración es el marco conceptual más notable en estos momentos en torno a la aplicación de la IA en el sector público. La investigación muestra una creciente externalización de tareas administrativas a la ciudadanía mediante autoservicios digitales, entre los que destaca la automatización robótica de procesos, las comunicaciones automatizadas, la clasificación de correos o mensajes electrónicos, o la búsqueda y recuperación documental. Veamos los casos detectados en la literatura.

En cuanto a la automatización robótica de procesos, se han documentado casos relevantes en países como Finlandia y Suecia. En Finlandia, se automatizaron aproximadamente medio millón de decisiones anuales relacionadas con ayudas y subsidios, aunque esta práctica cuestionada por parte del Defensor del Pueblo, debido a dudas sobre su base legal (Väänänen, 2021). En Suecia, por su parte, se ha implementado una automatización municipal a gran escala, que ha sido generalmente aceptada por la sociedad (Germundsson *et al.*, 2024; Ranerup y Svensson, 2024).

En el ámbito de la búsqueda documental, destaca el desarrollo del sistema GerPS-FIM-Microverse en Alemania, una representación semántica estandarizada para servicios públicos que ha facilitado la digitalización de más de 10 000 servicios administrativos y constituye un caso prometedor de

estandarización a gran escala (Raupach *et al.*, 2024). En relación con estos usos, se observan también aplicaciones recientes de modelos de lenguaje de gran tamaño³ basados en generación aumentada por recuperación (RAG).

En España, el Ayuntamiento de Barcelona, a través del Área de Derechos Sociales, ha implementado un sistema de apoyo a la toma de decisiones basado en inteligencia colectiva. Este sistema, conocido como DPR automático, realiza diagnósticos y provisiones de recursos a partir de un repositorio de entrevistas con profesionales. Está integrado en el sistema de información de los centros de servicios sociales y automatiza la codificación de demandas recibidas, problemas detectados y propuestas de prescripción de recursos (m4Social, 2024: 78).

Los sistemas de apoyo a decisiones (*decision-support systems*, DSS) han cobrado relevancia en los últimos años, especialmente en instituciones de seguridad social a nivel global (Germundsson *et al.*, 2024; Marienfeldt, 2024; Ruggia-Frick, 2021). En este contexto, se habla de toma de decisiones automática o automatizada (ADM) cuando el sistema posee cierta autonomía. Un ejemplo es el sistema RISK, en Dinamarca, diseñado para mejorar las evaluaciones de riesgo en protección infantil, aunque ha sido objeto de gran controversia (Ratner y Thylstrup, 2024).

En cuanto a la verificación de elegibilidad para prestaciones, el fondo de desempleo Töötukassa, en Estonia, ha implementado un sistema de toma de decisiones basado en reglas que automatiza aproximadamente la mitad de las decisiones, verificando automáticamente la información proporcionada por los solicitantes a través de bases de datos integradas (OCDE, 2024). Un sistema similar opera en el Departamento de Derechos Sociales de la Generalitat de Cataluña desde diciembre de 2018. Este motor de reglas verifica si un ciudadano o entidad cumple los criterios para recibir una prestación social. A fecha de enero de 2024, se habían desarrollado algoritmos para evaluar las necesidades básicas, los gastos del hogar, la pensión no contributiva de jubilación, el complemento de la citada pensión, y la prestación por nacimiento, acogida y adopción (m4Social, 2024: 90).

Una evolución reciente en estos sistemas es la incorporación de interfaces de toma de decisiones humano-IA. A diferencia de enfoques anteriores, se busca incluir al humano en el "bucle" del aprendizaje automático, de manera que se puedan trazar y contrastar los datos que conducen a un resultado, para así garantizar que este sea explicable desde una perspectiva humana. Estos sistemas combinan criterios algorítmicos con supervisión humana, como en el caso del robot Tengai, en Suecia, utilizado en

³ *Large language models* (LLM). También los hay pequeños (*small language models*), centrados en ámbitos de conocimiento específicos. El modelo de lenguaje de gran tamaño más conocido hoy en día es Chat GPT, de OpenAI.

procesos de reclutamiento en el empleo público. Se espera que estos procesos sean imparciales y estén mediados por la intervención humana (Centro Común de Investigación, 2020: tabla 11). Aunque estos sistemas puedan parecer similares a los de automatización robótica de procesos, se consideran cualitativamente distintos, ya que implican una interacción humana activa a lo largo del proceso algorítmico, especialmente en la toma de decisiones por parte de responsables públicos, técnicos o trabajadores sociales. En esta línea, en España se ha desarrollado wSocial, una herramienta basada en IA e impulsada por el Departamento de Derechos Sociales de la Generalitat de Cataluña. Utiliza palabras clave para detectar situaciones de vulnerabilidad y proponer intervenciones (Fundació iSocial, 2024).

En el ámbito de las comunicaciones automatizadas, el Instituto Nacional de la Seguridad Social de Italia ha implementado un sistema basado en IA para la clasificación y distribución de correos electrónicos certificados (*posta elettronica certificata*), sistema que permite su reenvío inmediato a las oficinas correspondientes (Centro Común de Investigación, 2024: 22). Finalmente, los chatbots y asistentes virtuales también se han incorporado en varios países. En Bélgica, el chatbot Ori, del servicio de empleo ONEM, responde preguntas sobre desempleo, bajas y ayudas. En Finlandia, el chatbot Kela-Kelpo/FPA-Folke asiste a la ciudadanía en la cumplimentación de solicitudes de prestaciones sociales en varios idiomas (OCDE, 2024: 22-23; Väänänen, 2021). En España, se están probando chatbots conversacionales para recomendar prestaciones, detectar necesidades en entrevistas, asistir en trámites y apoyar la gestión documental (m4Social, 2024: 91 y ss.).

3.2 Sistemas de perfilado y segmentación

Los sistemas de perfilado predictivo se consolidan como una de las aplicaciones emergentes de la inteligencia artificial en las políticas sociolaborales. Estos sistemas emplean algoritmos para clasificar a las personas usuarias en función de su probabilidad de requerir intervenciones específicas, lo que permite orientar recursos de manera más eficiente. Aunque el perfilado es, en parte, automático, esta automatización plantea riesgos significativos en términos de transparencia algorítmica. Para mitigarlos, se ha promovido el paradigma del *human-in-the-loop*, que introduce supervisión humana en el proceso de toma de decisiones.

Desde una perspectiva comparativa, los sistemas de perfilado presentan diferencias notables en cuanto a su diseño ético y orientación, lo que refleja la complejidad del asunto en términos éticos y normativos (Rachovitsa y Johann, 2022; Schmager *et al.*, 2024; Züger y Asghari, 2023). Para ilustrar esta diversidad, se presentan a continuación dos casos representativos de monitorización de beneficiarios —tanto actuales como potenciales— de servicios públicos en Austria y Finlandia.

En Austria, el Servicio Público de Empleo implementó el sistema Arbeitsmarkt-Chancen Assistenzsystem (AMAS), diseñado para calcular una puntuación de oportunidad de integración (*integration chance*) basada en los historiales laborales y las características personales de las personas solicitantes de empleo. Este sistema clasificaba a los individuos en tres categorías, según su probabilidad de reintegración en el mercado laboral (Achterhold *et al.*, 2025).

Por su parte, en Finlandia, el Centro de Pensiones desarrolló en 2018 un sistema predictivo de discapacidad laboral. Este algoritmo de aprendizaje automático utilizaba una técnica estadística de autoaprendizaje para prever si una persona se jubilaría con pensión por discapacidad en un plazo de dos años. El modelo, entrenado con datos socioeconómicos, de ingresos y prestaciones de 500 000 individuos, alcanzó una precisión del 78 % (Väänänen, 2021).

3.3 Sistemas de detección y respuesta temprana

En estrecha relación con los sistemas de perfilado y segmentación, hay un subconjunto relevante de aplicaciones orientadas a la detección y respuesta temprana ante necesidades o riesgos sociales. Estas herramientas buscan anticiparse a situaciones de vulnerabilidad mediante el análisis predictivo de datos, permitiendo intervenciones más ágiles y focalizadas. Seguidamente se presentan varios casos ilustrativos que reflejan la diversidad de enfoques y ámbitos de aplicación.

En el ámbito de la prestación de servicios, la ciudad de Gante (Bélgica) ha desarrollado un sistema que utiliza datos existentes para otorgar automáticamente descuentos en educación, cuidado infantil y otras ayudas a residentes en situación de vulnerabilidad (Kempeneer *et al.*, 2024).

En Eslovaquia, la Oficina Central de Trabajo, Asuntos Sociales y Familia (COLSAF) desarrolló, desde una perspectiva investigadora, un sistema de aprendizaje automático para predecir la duración del desempleo y optimizar el uso de recursos a dichos efectos, sistema que mostró una capacidad predictiva eficaz a partir del historial laboral, la educación (cualificaciones) y la edad (Gabrikova *et al.*, 2023). Cabe mencionar, conforme a lo puntualizado en el mismo estudio, que los modelos predictivos de tal sofisticación son susceptibles a los riesgos propios de las "cajas negras" algorítmicas, que detallaremos en el apartado 4.

En Dinamarca, varios proyectos algorítmicos han intentado detectar a la infancia en riesgo, incluyendo un controvertido sistema en el municipio de Gladsaxe, que experimentó con un sistema algorítmico para detectar niños y niñas en riesgo de abuso basándose en datos integrados desde diferentes sistemas, sociales, sanitarios y administrativos (Centro Común de Investigación, 2020). Este tipo de sistemas enfrentaron y enfrentan críticas considerables y fueron

judicializados, debido a su cuestionable enfoque y tratamiento de la privacidad (Cearns y Knox, 2024; Ratner y Schröder, 2024; Ratner y Thylstrup, 2024).

La predicción de la pobreza es un campo en el cual el aprendizaje automático y la ciencia de datos tendrían mucho que ofrecer; actualmente, existe literatura sobre imágenes satelitales que incluyen pautas para su aplicación en políticas del desarrollo (Hall *et al.*, 2023), pero también aplicaciones más directas y cercanas, por ejemplo, en pobreza energética. Así, en España, la Administración Abierta de Cataluña (AOC) ha puesto en marcha un servicio de automatización para la elaboración de informes de pobreza energética. Anteriormente, estos informes se redactaban a mano, lo que los hacía ineficientes y propensos a errores. Se ha implementado una plataforma en la nube que “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” (m4Social, 2024: 103). Con un enfoque más orientado a la investigación, también se desarrolló un marco de clasificación de los hogares en los Países Bajos, según cuatro categorías de riesgo de pobreza energética. Este sistema utiliza algoritmos avanzados para identificar factores predictivos, como el valor de la vivienda y su estatus de propiedad, la antigüedad, el número de personas por hogar y la densidad de población (Dalla Longa *et al.*, 2021).

3.4. Participación ciudadana

La convergencia de IA, datos masivos y plataformas digitales impulsa nuevas formas de colaboración ciudadana. Las redes sociales dedicadas tienen el potencial de acercar a evaluadores, gestores de programas y usuarios finales, creando espacios de interacción antes inexistentes, propiciando el acceso a aplicaciones informáticas y móviles de fácil manejo y fomentando la participación de las comunidades tanto en la creación como en la difusión del contenido evaluativo (Picciotto, 2020).

En el contexto europeo, existen ya ejemplos concretos de esta integración tecnológica en procesos participativos. Una reciente revisión de la literatura (Babšek *et al.*, 2025) documenta casos como CitizenLab en Bélgica o Civocracy en Alemania. CitizenLab utiliza *machine learning* e IA generativa para procesar ideas del público y transformarlas en recomendaciones aplicables a iniciativas ambientales. Por su parte, Civocracy fomenta la participación comunitaria mediante discusiones transparentes, colaboración estructurada y análisis de sentimiento, permitiendo una evaluación continua de la opinión pública.

Además, la evaluación participativa puede ayudar a mitigar los efectos adversos del *big data* y la IA, guiando la gobernanza de manera informada: los procesos de evaluación participativos pueden ser un vehículo para garantizar una toma de decisiones ética y abordar

los riesgos sociales asociados a las aplicaciones de macrodatos, promoviendo la transparencia y la rendición de cuentas (Picciotto, 2020).

4. Retos técnicos, sociales y éticos

La literatura revisada muestra cómo la aplicación de IA ha producido un incremento de eficiencia; una reducción de costes y tiempo (m4Social, 2024; OCDE, 2024; Perron *et al.*, 2024; Ranerup y Svensson, 2024; Väänänen, 2021), de procesos y asignación de recursos (Marienfeldt, 2024; OCDE, 2024; Schmager *et al.*, 2024; Väänänen, 2021), e incluso mejoras en calidad, accesibilidad y personalización de servicios (Lee-Archer, 2023; m4Social, 2024). Además, la IA nos permite —en cierta medida— la prestación proactiva de servicios y la intervención temprana (Kempeneer *et al.*, 2024; Lehtiniemi, 2024; m4Social, 2024; OCDE, 2024; Ratner y Schröder, 2024) y, con las salvaguardas adecuadas, podría fortalecer el apoyo específico a grupos vulnerables (m4Social, 2024; Valle Escolano, 2023).

Pueden servir como caso de éxito, por ejemplo, los sistemas de interoperabilidad e intercambio de datos entre instituciones de la seguridad social para mejorar la prestación de servicios a la ciudadanía y a la patronal en Bélgica (Lee-Archer, 2023) o de arquitectura digital interconectada en las aplicaciones de inteligencia artificial y toma de decisiones automatizada en la Seguridad Social finlandesa (Väänänen, 2021) —muy dependiente, a su vez, de la confianza de la ciudadanía en el Gobierno—. Estos casos, junto con los estudiados en Cataluña, muestran las posibilidades en materia de interoperabilidad de datos, pero a la vez resaltan las dificultades que existen para analizar de forma coordinada y compartir, no ya datos, sino tecnologías.

Sin embargo, estas mismas ventajas implican retos técnicos, sociales y éticos. La evidencia documental revela una relación compleja entre eficiencia y otros valores públicos, particularmente en la eficiencia administrativa y la necesidad de una adopción progresiva e incremental frente a la disrupción innovadora (Minguíjon y Serrano-Martínez, 2022). En lo referente a la privacidad y la no discriminación, se observa un cuestionamiento sobre la proporcionalidad de los datos a lo largo de la literatura. La opacidad algorítmica en sí misma y los derechos fundamentales también aparecen en una relación conflictiva (Rachovitsa y Johann, 2022; Schmager *et al.*, 2024; Züger y Asghari, 2023) y se plantean problemas como la atribución de responsabilidad (riesgo moral), el sesgo algorítmico (riesgo de amplificar sesgos históricos y riesgo preocupante de sesgo étnico en detección de fraude), las cajas negras (falta de transparencia), la fragmentación de datos y los silos, y la incompatibilidad e interoperabilidad.

Efecto de riesgo moral

Se ha observado que, en ocasiones, los funcionarios públicos abdican de su responsabilidad en la toma de decisiones automatizadas. Se cita el caso de

la autoridad fiscal holandesa (Belastingdienst), que implementó de 2013 a 2021 un algoritmo de aprendizaje automático para procesar reclamaciones de subsidios para guardería infantil, el cual señalaba incorrectamente a muchas familias como potenciales defraudadoras, un error que afectó especialmente a personas de bajos ingresos y de minorías étnicas. El caso ha evidenciado cómo los empleados simplemente aceptaban las decisiones del algoritmo, sin cuestionar sus resultados: los funcionarios "se despojaron de responsabilidad moral y legal" al aceptar los resultados de la máquina, lo que llevó a clasificar incorrectamente a muchas familias como posibles defraudadoras y dio pie al efecto de riesgo moral (Lee-Archer, 2023).

Sesgo algorítmico

Cuando los sistemas de IA utilizan datos históricos, existe el riesgo de importar y amplificar sesgos, de los cuales las personas usuarias podrían no ser conscientes (Centro Común de Investigación, 2020). Esto es, entre otras cosas, el sesgo algorítmico. De encontrarse, por ejemplo, un sesgo étnico en el modelo, deberían poder identificarse las relaciones que producían ese sesgo y tratar de reducirlas (Centro Común de Investigación, 2024; Rachovitsa y Johann, 2022; Ratner y Thylstrup, 2024; Schmager *et al.*, 2024; Valle Escolano, 2023). Existen problemas inherentes a la dependencia de modelos predictivos, los cuales reconfiguran a las personas como colecciones de rasgos y eventos pasados: ello es una base potencial para sesgos sistémicos (Lehtiniemi, 2024).

Cajas negras

Los sistemas de IA en el sector público a menudo operan como cajas negras, lo cual dificulta entender cómo se toman las decisiones automatizadas. Esta opacidad puede erosionar la confianza y limitar la rendición de cuentas (Centro Común de Investigación, 2024). El sistema AMAS, utilizado por el Servicio Público de Empleo de Austria, es un ejemplo de ello. Inicialmente, ni los datos utilizados ni el modelo completo fueron divulgados, lo que dificultaba el escrutinio externo. Solo tras presiones se publicó una representación basada en regresión logística, que reveló que atributos como el género y la nacionalidad incidían negativamente en la puntuación de reintegración laboral y actuaban como un sesgo algorítmico. Esta opacidad suscitó preocupación sobre la reproducción de discriminaciones históricas, haciendo de AMAS un ejemplo paradigmático de "caja negra algorítmica." Frente a ello, indicadores de equidad⁴ y técnicas de mitigación de sesgos⁵

⁴ Medidas estadísticas que evalúan la equidad de un modelo algorítmico, observando cómo se distribuyen los resultados y errores entre distintos grupos definidos por atributos sensibles, como el género. Estas métricas permiten detectar disparidades sistemáticas en las decisiones automatizadas.

⁵ Técnicas de mitigación de sesgos como: reponderación de datos, representaciones justas, regresión logística con restricciones de equidad y posprocesamiento por paridad de oportunidades.

pueden permitir abrir y corregir decisiones algorítmicas opacas, como ha demostrado finalmente este caso (Achterhold *et al.*, 2025). En este sentido, la adopción de IA debe preservar valores públicos como la igualdad, la neutralidad y la imparcialidad, propios del mandato del servicio público, así como la transparencia, la rendición de cuentas y la posibilidad de objeción, especialmente en decisiones que afectan directamente la vida de la ciudadanía (Schmager *et al.*, 2024).

Fragmentación de datos y silos organizacionales

En múltiples casos, se ha observado cómo los departamentos o niveles administrativos dentro de una misma organización manejan sus propios sistemas de información y prácticas de gestión de datos, sin una coordinación efectiva. Esta falta de interoperabilidad impide compartir información de forma fluida y coherente, lo que afecta negativamente la calidad de las decisiones automatizadas y limita el potencial de los sistemas de IA. La fragmentación de datos también complica el entrenamiento de algoritmos con información representativa y de calidad, y aumenta el riesgo de decisiones sesgadas (Centro Común de Investigación, 2024).

Incompatibilidad e interoperabilidad

En un plano más general, la imagen en conjunto de los servicios públicos electrónicos sigue siendo algo fragmentada, debido al gran número de instituciones involucradas (Lee-Archer, 2023). Hoy en día existe una necesidad de infraestructuras integradas que raramente están disponibles, pero, además, hay tensión entre el potencial analítico de la integración de datos y las restricciones de privacidad y propósito específico (Schmager *et al.*, 2024). También se pueden observar casos de desalineación entre objetivos de proveedores tecnológicos, administraciones públicas y ciudadanía (Cearns y Knox, 2024).

5. Conclusiones

Esta revisión exploratoria ha permitido mapear el estado actual de la aplicación de la IA en la implementación de políticas públicas de empleo e inclusión social en el contexto europeo, revelando una progresión tecnológica que transita desde funciones administrativas básicas —como la automatización robótica de procesos, la clasificación de comunicaciones o la gestión documental— hacia sistemas algorítmicos más sofisticados orientados al apoyo en la toma de decisiones, que incorporan formalmente el paradigma *human-in-the-loop* como mecanismo de supervisión.

El ecosistema tecnológico detectado se nutre principalmente del aprendizaje automático en sus variantes supervisada y no supervisada, complementado con el procesamiento del lenguaje natural, los modelos generativos y el análisis visual. Los casos analizados documentan beneficios

operativos tangibles en términos de eficiencia, reducción de tiempos y costes, optimización en la asignación de recursos, y mejoras en accesibilidad y personalización de servicios.

No obstante, la revisión pone de manifiesto tensiones estructurales entre los objetivos de eficiencia administrativa y otros valores públicos fundamentales. Los casos analizados revelan problemas recurrentes: el efecto de riesgo moral en la toma de decisiones automatizadas; el sesgo algorítmico, que reproduce discriminaciones históricas; la opacidad de los sistemas, que dificulta la rendición de cuentas, y la fragmentación organizacional, que limita la interoperabilidad de datos. Estos desafíos adquieren particular relevancia cuando afectan a poblaciones

vulnerables, donde las consecuencias de decisiones algorítmicas erróneas o sesgadas pueden perpetuar o amplificar desigualdades preexistentes.

La contribución de este estudio radica en proporcionar un mapeo sistemático que permite comprender tanto las potencialidades como las limitaciones actuales de estas tecnologías en un ámbito especialmente sensible de la acción pública. La evidencia analizada subraya que la implementación efectiva y ética de la IA en políticas de empleo e inclusión social requiere no solo capacidad técnica, sino marcos de gobernanza robustos que garanticen transparencia, equidad y salvaguarda de derechos fundamentales.

- ACHTERHOLD, E.; MÜHLBÖCK, M.; STEIBER, N. y KERN, C. (2025): "Fairness in algorithmic profiling: the AMAS case", *Minds and Machines*, vol. 35, n.º 1, art. 9, <<https://doi.org/10.1007/s11023-024-09706-9>>.
- ASOCIACIÓN ASLAN (2025): "InfoSIA: inteligencia artificial para la asistencia a profesionales en la propuesta de prestaciones y recursos idóneos a las personas usuarias de servicios sociales" [candidatura presentada a la XVII Convocatoria Premios Transformación Digital], Madrid, ASLAN, <<https://aslan.es/mejora-de-la-eficiencia-y-experiencia-del-empleado-candidatura2025/>>.
- BABŠEK, M.; RAVŠELJ, D.; UMEK, L. y ARISTOVNIK, A. (2025): "Artificial intelligence adoption in public administration: an overview of top-cited articles and practical applications", *AI*, vol. 6, n.º 3, art. 44, <<https://doi.org/10.3390/ai6030044>>.
- BARREDO, A. *et al.* (2020): "Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI", *Information Fusion*, vol. 58, págs. 82-115, <<https://doi.org/10.1016/j.inffus.2019.12.012>>.
- BRIOSCÚ, A.; LAURINGSON, A.; SAINT-MARTIN, A. y XENOGIANI, T. (2024): *A new dawn for public employment services: service delivery in the age of artificial intelligence*, serie OECD Artificial Intelligence Papers, n.º 19, París, OECD Publishing, <<https://doi.org/10.1787/5dc3eb8e-en>>.
- CARABANTES, M. (2020): "Black-box artificial intelligence: an epistemological and critical analysis", *AI & Society*, vol. 35, n.º 2, págs. 309-317, <<https://doi.org/10.1007/s00146-019-00888-w>>.
- CEARNS, J. y KNOX, H. (2024): "The data consensus and the public good in children's social services", *The Cambridge Journal of Anthropology*, vol. 42, n.º 1, págs. 23-41, <<https://doi.org/10.3167/cja.2024.420103>>.
- CENTRO COMÚN DE INVESTIGACIÓN (2020): *AI Watch - artificial intelligence in public services*, Bruselas, Comisión Europea, <https://ai-watch.ec.europa.eu/publications/ai-watch-artificial-intelligence-public-services_en>.
- (2024): *Competencies and governance practices for AI in the public sector*, Luxemburgo, Oficina de Publicaciones de la Unión Europea, <<https://doi.org/10.2760/7895569>>.
- COMISIÓN EUROPEA (2018): *Inteligencia artificial para Europa*, COM(2018) 237 final, Bruselas, Comisión Europea, <<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=COM:2018:237:FIN>>.
- (2024): *Adoption of AI, blockchain and other emerging technologies within the European public sector: a public sector tech watch report*. Luxemburgo, Oficina de Publicaciones de la Unión Europea, <<https://doi.org/10.2799/3438251>>.
- DALLA LONGA, F.; SWEERTS, B. y VAN DER ZWAAN, B. (2021): "Exploring the complex origins of energy poverty in The Netherlands with machine learning", *Energy Policy*, vol. 156, art. 112373, <<https://doi.org/10.1016/j.enpol.2021.112373>>.
- DIRECCIÓN GENERAL DE EMPLEO, ASUNTOS SOCIALES E INCLUSIÓN e ICF (2025): *Opportunities of AI within PES processes and services: Exploring PES experiences, best practices and emerging business value*. Luxemburgo, Oficina de Publicaciones de la Unión Europea, <<https://doi.org/10.2767/84293>>.
- FRESNO CONSULTING (2025a): *Exploración, análisis y prospección en la aplicación de la IA en los procesos de evaluación de las políticas públicas. Casos de uso*, s. l., Besaldi, <<https://euskadi.eus/>>

- contenidos/informacion/besaldi_documentos/es_def/adjuntos/Maquetado-Besaldi_Estudio_IA_Informe_vf.pdf>.
- FRESNO CONSULTING (2025b): *Exploración, análisis y prospección en la aplicación de la IA en los procesos de evaluación de las políticas públicas. Informe de resultados*, s. l., Besaldi, <https://euskadi.eus/contenidos/informacion/besaldi_documentos/es_def/adjuntos/Maquetado-Besaldi_Estudio_IA_Informe_vf.pdf>.
- FUNDACIÓ ISOCIAL (2024): "La inteligencia artificial en los servicios sociales: análisis predictiva e identificación de necesidades de intervención", Barcelona, Fundació iSocial, <<https://isocial.cat/es/la-inteligencia-artificial-en-los-servicios-sociales-analisis-predictiva-e-identificacion-de-necesidades-de-intervencion/>>.
- GABRIKOVA, B.; SVABOVA, L. y KRAMAROVA, K. (2023): "Machine learning ensemble modelling for predicting unemployment duration", *Applied Sciences*, vol. 13, n.º 18, art. 18, <<https://doi.org/10.3390/app131810146>>.
- GERMUNDSSON, N.; STRANZ, H. y BERGMARK, Å. (2024): "Reducing administration? Examining the alignment of robotic process automation and social assistance in Swedish personal social services", *Nordic Social Work Research*, <<https://doi.org/10.1080/2156857X.2024.2440720>>.
- GOODFELLOW, I.; COURVILLE, A. y BENGIO, Y. (2016): *Deep learning*, Cambridge, The MIT Press.
- HALL, O.; DOMPAE, F.; WAHAB, I. y DZANKU, F. M. (2023): "A review of machine learning and satellite imagery for poverty prediction: implications for development research and applications", *Journal of International Development*, vol. 35, n.º 7, págs. 1753-1768, <<https://doi.org/10.1002/jid.3751>>.
- ISO e IEC (2022): *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*, ISO/IEC 22989:2022(E), Ginebra, International Organization for Standardization e International Electrotechnical Commission.
- KEMPENEER, S.; RANCHORDAS, S. y VAN DE WETERING, S. (2024): "AI failure, AI success, and AI power dynamics in the public sector", *SSRN*, <<https://doi.org/10.2139/ssrn.4983622>>.
- LEE-ARCHER, B. (2023): *Effects of digitalization on the human centrality of social security administration and services*, serie ILO Working Papers, n.º 87, Ginebra, Organización Internacional del Trabajo, <<https://doi.org/10.54394/PMPD3825>>.
- LEHTINIEMI, T. (2024): "Contextual social valences for artificial intelligence: anticipation that matters in social work", *Information, Communication & Society*, vol. 27, n.º 6, págs. 1110-1125, <<https://doi.org/10.1080/1369118X.2023.2234987>>.
- LEVAC, D.; COLQUHOUN, H. y O'BRIEN, K. K. (2010): "Scoping studies: advancing the methodology", *Implementation Science*, vol. 5, art. 69, <<https://doi.org/10.1186/1748-5908-5-69>>.
- M4SOCIAL (2024): *Radars de algoritmos de IA y procesos de decisión automatizada para el acceso a los derechos sociales en Cataluña*, Barcelona, m4Social, <<https://m4social.org/es/recursos/radar-dalgoritmes-dia-i-processos-de-decisio-automatitzada-per-a-lacces-als-drets-socials-a-catalunya/>>.
- MARIENFELDT, J. (2024): "Does digital government hollow out the essence of street-level bureaucracy? A systematic literature review of how digital tools foster curtailment, enablement and continuation of street-level decision-making", *Social Policy & Administration*, vol. 58, n.º 5, págs. 831-855, <<https://doi.org/10.1111/spol.12991>>.
- MINGUIJON, J. y SERRANO-MARTINEZ, C. (2022): "La inteligencia artificial en los servicios sociales: estado de la cuestión y posibles desarrollos futuros", *Cuadernos de Trabajo Social*, vol. 35, n.º 2, art. 2, <<https://doi.org/10.5209/cuts.78747>>.
- MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA (2024): *Estrategia de Inteligencia Artificial 2024*, Madrid, Ministerio para la Transformación Digital y de la Función Pública, <https://portal.mineco.gob.es/es-es/digitalizacion/IA/Documents/Estrategia_IA_2024.pdf>.
- MOYANO-ARIAS, R. J.; SALAZAR-ÁLVAREZ, E. G. y TOALOMBO-VARGAS, V. M. (2024): "Matemáticas aplicadas a la programación: una revisión sobre la solución de algoritmos complejos", *MQRInvestigar*, vol. 8, n.º 4, págs. 3667-3692, <<https://doi.org/10.56048/MQR20225.8.4.2024.3667-3692>>.
- OCDE (2019): *Recommendation of the Council on artificial intelligence*, OECD/LEGAL/0449, París, OECD Publishing, <<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>>.
- (2024): *Using AI to manage minimum income benefits and unemployment assistance: opportunities, risks and possible policy directions*, serie OECD Artificial Intelligence Papers, n.º 21, París, OECD Publishing, <<https://doi.org/10.1787/718c93a1-en>>.
- (2025): *Governing with artificial intelligence*. París, OECD Publishing, <https://www.oecd.org/en/publications/governing-with-artificial-intelligence_795de142-en.html>.
- PERRON, B. E.; HILTZ, B. S.; KHANG, E. M. y SAVAS, S. A. (2025): "AI-enhanced social work: developing and evaluating retrieval-augmented generation (RAG) support systems", *Journal of Social Work Education*, vol. 61, n.º 1, págs. 3-13, <<https://doi.org/10.1080/10437797.2024.2411172>>.
- PERRON, B. E.; LUAN, H.; VICTOR, B. G.; HILTZ-PERRON, O. y RYAN, J. (2024): "Moving beyond ChatGPT: local large language models (LLMs) and the secure analysis of confidential unstructured text data in social work research", *Research on Social Work Practice*, vol. 35, n.º 6, págs. 695-710, <<https://doi.org/10.1177/10497315241280686>>.
- PICCIOTTO, R. (2020): "Evaluation and the Big Data challenge", *American Journal of Evaluation*, vol. 41, n.º 2, págs. 166-181, <<https://doi.org/10.1177/1098214019850334>>.
- RACHOVITSA, A. y JOHANN, N. (2022): "The human rights implications of the use of AI in the digital welfare state: lessons learned from the Dutch SyRI case", *Human Rights Law Review*, vol. 22, n.º 2, art. ngac010, <<https://doi.org/10.1093/hrlr/ngac010>>.
- RANERUP, A. y SVENSSON, L. (2024): "Value positions in the implementation of automated decision-making in social assistance", *Nordic Social Work Research*, vol. 14, n.º 4, págs. 471-485, <<https://doi.org/10.1080/2156857X.2022.2062040>>.

- RATNER, H. F. y SCHRØDER, I. (2024): "Ethical plateaus in Danish child protection services: the rise and demise of algorithmic models", *Science & Technology Studies*, vol. 37, n.º 3, art. 3, <<https://doi.org/10.23987/sts.126011>>.
- RATNER, H. F. y THYLSTRUP, N. B. (2024): "Citizens' data afterlives: practices of dataset inclusion in machine learning for public welfare", *AI & Society*, vol. 40, págs. 1183-1193, <<https://doi.org/10.1007/s00146-024-01920-4>>.
- RAUPACH, M.; ENDERLING, M.; FEDDOUL, L.; LEGNER, H.; MAUCH, M. y KÖNIG-RIES, B. (2024): "Towards a semantic format for FIM: supporting German public services using the GerPS-FIM-Microverse ontology pipeline" [comunicación en congreso], en ASGHARI, H. y ZÜGER, T. (eds.), *2nd Workshop on 'Public Interest AI' co-located with the 47th German Conference on AI (KI 2024)*, 23-09-2024, Würzburg, Julius-Maximilians Universität, <<https://ceur-ws.org/Vol-3958/piai24-paper3.pdf>>.
- RAYA DÍEZ, E.; TRUJILLO CARMONA, M. y CARBONERO MUÑOZ, D. (2021): "Using Big Data to manage social inclusion programs", *The Journal of Sociology & Social Welfare*, vol. 48, n.º 3, <<https://doi.org/10.15453/0191-5096.4556>>.
- RUGGIA-FRICK, R. (2021): "Applying emerging data-driven technologies in social security. Country experiences and ISSA guidelines", *Ubezpieczenia Społeczne. Teoria i Praktyka*, vol. 150, n.º 3, <<https://doi.org/10.5604/01.3001.0015.5230>>.
- SAURA, J. R. y DEBASA, F. (eds.) (2022): *Handbook of research on artificial intelligence in government practices and processes*. Hershey, IGI Global, <<https://doi.org/10.4018/978-1-7998-9609-8>>.
- SCHMAGER, S.; GRØDER, C. H.; PARMIGGIANI, E.; PAPPAS, I. y VASSILAKOPOULOU, P. (2024): "Exploring citizens' stances on AI in public services: a social contract perspective", *Data & Policy*, vol. 6, art. e19, <<https://doi.org/10.1017/dap.2024.13>>.
- SERVICIO PÚBLICO DE EMPLEO ESTATAL (2024): *La IA aplicada a los servicios públicos de empleo. Hacia un sistema de intermediación, orientación y formación basado en competencias profesionales*. Madrid, Servicio Público de Empleo Estatal, <<https://www.sepe.es/HomeSepe/que-es-observatorio/Hipatia/cuadernos-mercado-trabajo/revista-cuadernos-mercado-trabajo/detalle-articulo.html?detail=/revista/Econom-a-Social/laiaplicadaalosserviciospublicosdeempleohaciaunsistemadeintermediacionorientacionyformacionbasadoencompetenciasprofesionales>>.
- SUBIRATS, J. (ed.) (2008): *Análisis y gestión de políticas públicas*, 1.ª ed., Barcelona, Ariel.
- TANGI, L.; VAN NOORDT C.; COMBETTO, M.; GATTWINKEL, D. y PIGNATELLI, F. (2022): *AI watch. European landscape on the use of artificial intelligence by the public sector*, EUR 31088 EN, Luxemburgo, Oficina de Publicaciones de la Unión Europea, <<https://doi.org/10.2760/39336>>.
- VÄÄNÄNEN, N. (2021): "The digital transition of social security in Finland. Frontrunner experiencing headwinds?", *Ubezpieczenia Społeczne. Teoria i Praktyka*, vol. 4, págs. 71-86, <<https://doi.org/10.5604/01.3001.0015.5251>>.
- VALLE ESCOLANO, R. (2023): "Inteligencia artificial y derechos de las personas con discapacidad: el poder de los algoritmos", *Revista Española de Discapacidad*, vol. 11, n.º 1, págs. 7-28, <<https://doi.org/doi.org/10.5569/2340-5104.11.01.01>>.
- ZÜGER, T. y ASGHARI, H. (2023): "AI for the public. How public interest theory shifts the discourse on AI", *AI & Society*, vol. 38, n.º 2, págs. 815-828, <<https://doi.org/10.1007/s00146-022-01480-5>>.

Anexo I. Resumen de hallazgos en casos de uso

	Caso de uso	Tecnología usada	Riesgos y limitaciones
Achterhold <i>et al.</i> (2025)	El sistema AMAS, del Servicio Público de Empleo de Austria, predice probabilidades de reintegración laboral de jóvenes desempleados para asignar recursos de apoyo, clasificándolos según riesgo de desempleo prolongado.	Evaluación empírica de equidad y aplicación de técnicas de mitigación de sesgos.	Limitación sociodemográfica del grupo (jóvenes), medidas de apoyo insuficientemente definibles, limitación severa en métricas de oportunidad y en equidad.
Cearns y Knox (2024)	Sistema para detectar niñas y niños en riesgo que necesitan protección en servicios sociales del Reino Unido. Explora el "consenso de datos" entre trabajadores sociales y científicos de datos.	Sistema de ML predictivo para detección de vulnerabilidad infantil.	Conflictos de interpretación de datos; disparidad de motivaciones (ayuntamiento/proveedor); preocupación ética (<i>profiling</i>); tensión humano vs. dato; la definición de "dato" varía.
Centro Común de Investigación (2024)	Desarrollo de competencias y prácticas de gobernanza para adoptar la IA en organizaciones del sector público europeo de forma ética y efectiva.	Chatbots, NLP, visión artificial, ML predictivo, detección de fraude algorítmica.	Financiación; escasez de talento; silos/preparación de datos (80% trabajo); interpretación de reglas; falta de transparencia de los algoritmos.
Centro Común de Investigación (2020)	Mapeo del uso de IA en servicios públicos en la UE: mejora servicios, diseño políticas, gestión interna, atención ciudadana mediante chatbots y toma de decisiones algorítmicas automatizadas.	Diversas aplicaciones de IA (correo electrónico, chatbots, predicción crimen/precios/personal).	Retos en protección de privacidad, ética algorítmica, protección del trabajo y transparencia algorítmica.
Dalla Longa <i>et al.</i> (2021)	Predicción de riesgo de pobreza energética en hogares de los Países Bajos mediante clasificación basada en ingreso y gasto energético.	Clasificador ML con parámetros socioeconómicos.	Dificultad de detectar hogares vulnerables subrepresentados; necesita datos grandes/representativos; mecanismos causales complejos; la elección umbrales de ML afecta al rendimiento.
Fundació iSocial (2024)	Detección de vulnerabilidad social y análisis predictivo en servicios sociales para optimizar recursos y facilitar intervención anticipada personalizada.	Análisis de texto con palabras clave, modelos predictivos ML, NLP, visualización de datos y tendencias.	No se señalan.
Gabrikova <i>et al.</i> (2023)	Predicción de la duración del desempleo en Eslovaquia, categorizando a las personas en cuatro grupos temporales, según características individuales.	Uso combinado de ML, CART, CHAID, análisis discriminante, regresión logística y boosting para el balanceo de datos.	Caja negra; necesidad de gobernanza de datos; actualización de datos constante.
Germundsson <i>et al.</i> (2024)	Automatización de tareas administrativas en atención social para personas vulnerables en cuatro municipios suecos, analizando el impacto en la práctica profesional.	RPA basada en reglas predeterminadas, algoritmos estructurados para transferencia y cálculo de datos.	El sistema se detiene ante errores menores en la introducción de datos. Éticamente, la estandarización requerida por el RPA entra en conflicto con la necesidad de evaluaciones individualizadas exigidas por ley.
Kempeneer <i>et al.</i> (2024)	IA en servicios sociales/fraude; impacto desproporcionado en personas vulnerables. Ejemplos: servicio proactivo (BEL), fraude (DNK), discriminación (NLD), datos malos (UK), <i>chatbot</i> gradual (UK)	Marco conceptual, sin especificar tecnologías concretas de implementación.	Mala calidad datos (ignorada), sesgo/racismo institucional (NLD), eficiencia > equidad, incumplimiento del RGPD (frecuente).
Lee-Archer (2023)	Digitalización de la administración de seguridad social para servicios centrados en personas, integrando datos entre agencias, automatizando decisiones y mejorando detección de fraude con supervisión humana.	ML, <i>big data analytics</i> , RPA, <i>blockchain</i> , biometría, plataformas móviles, chatbots, API.	Sesgo algorítmico, exclusión digital, falta de transparencia, violación de privacidad, falsos positivos, erosión confianza pública, fragmentación de servicios.
Lehtiniemi (2024)	IA predictiva en bienestar infantil finlandés para identificar riesgos de colocación de emergencia o custodia, mediante el análisis de historiales sociosanitarios familiares completos.	ML predictivo, bases de datos administrativas combinadas, registros electrónicos salud/servicios sociales, interfaz integrada sistema información.	Reducción de personas a características, sesgo histórico, omisión de factores protectores, predicciones descontextualizadas, estigmatización de clientes, daño en las relaciones profesionales.
m4Social (2024)	Sistemas de IA implementados por administraciones públicas catalanas para gestionar el acceso a derechos sociales y automatizar tareas administrativas.	ML, NLP, chatbots, RPA, reconocimiento facial, algoritmos de clasificación, motores de reglas automatizados.	Sistemas internos con garantías limitadas, cautela, explicabilidad, necesidad de monitorización continua y evaluación del impacto real.
Marienfeldt (2024)	Revisión sistemática sobre el impacto de herramientas digitales en el trabajo de profesionales de primera línea en servicios públicos.	Sistemas de gestión de casos, evaluación de riesgos, decisiones automatizadas y portales de autoservicio digital.	Diseño rígido (ignora la complejidad social/individualidad); aplicación inflexible de reglas universales.
Minguijon y Serrano-Martinez (2022)	Análisis del grado de adaptación de los servicios sociales españoles a la IA y propuesta de modelo para integrar la IA en diferentes fases de intervención social.	Modelo que evalúa la IA en distintas fases de intervención y grados de automatización.	Necesidad de un silo de datos, de participación de profesionales del trabajo social y de una implementación gradual. Son claves la interoperabilidad y el apoyo público decidido.

	Caso de uso	Tecnología usada	Riesgos y limitaciones
OCDE (2024)	IA para gestionar prestaciones condicionadas (ingreso mínimo y ayuda al desempleo): para informar, tramitar, evaluar la elegibilidad y detectar pagos indebidos, así como para aumentar el acceso y la eficiencia.	Chatbots y asistentes (NLP), aprendizaje automático, minería de datos y ADM con reglas/datos interinstitucionales.	Sesgos y errores en elegibilidad, privacidad frágil, falta de transparencia/explicabilidad, responsabilidades difusas, brecha digital.
Perron <i>et al.</i> (2024)	Analizar de forma segura textos confidenciales para detectar y extraer problemas de sustancias en expedientes infantiles.	LLM locales (Mistral-7B, Mixtral-8x7B, Llama-3 8B/70B); <i>zero-shot</i> ; clasificación/extracción de 2956 resúmenes.	Prompts no sistemáticos (posible sesgo de Llama3), obsolescencia rápida de los resultados, errores en "verdad campo", restricciones del LLM propietario en cuestiones sensibles.
Perron <i>et al.</i> (2025)	Desarrollo y evaluación de sistemas RAG para apoyar decisiones y atender a clientes en trabajo social, integrando bases de conocimiento institucional con modelos generativos.	RAG con LLM + recuperación documental, orígenes de datos organizacionales; evaluación de la precisión/ fiabilidad.	Riesgo de alucinaciones y sesgos, depende de la calidad del repositorio, privacidad, generalización aún limitada.
Rachovitsa y Johann (2022)	Evaluación del impacto de los algoritmos antifraude en el bienestar digital, usando el caso SyRI para orientar salvaguardas y estándares.	Algoritmos de perfilado de riesgos (SyRI); análisis jurídico comparado; estándares CEDH/ONU.	Opacidad deliberada, sin notificación a afectados, riesgo de discriminación (<i>targeting</i>), datos amplios, obstaculiza la revisión judicial.
Ranerup y Svensson (2024)	Explora cómo distintas prioridades normativas plasman el diseño/uso de decisiones automatizadas en la atención social en cuatro municipios suecos.	Sistemas ADM para prestaciones; análisis comparado y participativo con actores municipales.	Metodología centrada en intenciones, sin desagregar grupos; valores profesionales divergentes; problemas técnicos y de tiempo.
Ratner y Schröder (2024)	Comparación de cuatro modelos de predicción del riesgo de maltrato infantil (ML sobre datos históricos administrativos).	Modelos predictivos de riesgo, aprendizaje automático, sistemas de apoyo en decisiones.	Restricciones legales de la fusión de datos, riesgo de sobreinformación sin base legal, necesidad de refinar datos vs. sesgos detectados (ej.: étnico), proyectos cancelados o limitados
Raupach <i>et al.</i> (2024)	Unificar y digitalizar trámites administrativos, mejorando la búsqueda, la interoperabilidad y la tramitación integral con representación semántica de procesos y formularios.	Ontologías y grafos de conocimiento, BPMN, microservicios; API, plantillas FIM, mapeo XProzess/ XDatenfeld.	No claramente reportados.
Saura y Debasa (2022)	Guía práctica para implantar la IA en administraciones públicas: diseñar políticas y servicios (educación, seguridad, trámites), con casos y marcos de gobernanza.	ML, NLP, algoritmos de clasificación y recomendación, minería de datos, motores de reglas, analítica predictiva.	Dependencia de datos secundarios, posible sesgo de publicación (éxitos), generalización limitada.
Schmager <i>et al.</i> (2024)	Prototipo IA que predice la duración de la baja médica y apoya informativamente a los gestores/trabajadores sociales, diseñado y evaluado bajo las "lentes del contrato social."	ML, algoritmos de clasificación, sistemas de apoyo en decisiones, explicabilidad/XAI, trazabilidad de datos.	Muestra pequeña (20) y contexto noruego, prototipo no implementado, sesgos de confianza y deseabilidad.
Väänänen (2021)	Evalúa la digitalización del sistema de seguridad social finlandés, identificando retos legales y éticos del uso de la IA y la decisión automatizada.	Plataformas digitales, automatización de procesos, ADM.	Legalidad RPA/ADM cuestionada (Defensor Pueblo: "decisiones no solo automáticas"), vacío legal vs. Constitución, nueva legislación necesaria/urgente según instituciones.
Valle Escolano (2023)	Revisión sobre cómo la IA afecta los derechos de personas con discapacidad: beneficios, sesgos y salvaguardas jurídicas (CDPD), con recomendaciones para uso inclusivo.	ML, <i>big data</i> , reconocimiento facial/emocional, perfilado automatizado, sistemas ADM.	Equipos diversos, transparencia, involucrar a usuarios; gobernanza ética, regulación específica, combatir el sesgo; investigar la inclusión, la IA como innovación social.

ADM: *automated decision-making* (toma de decisiones automatizada).

API: *application programming interface* (interfaz de programación de aplicaciones).

BPMN: *business process model and notation* (modelo y notación de procesos de negocio).

CART: *decision tree learning* (aprendizaje basado en árboles de decisión).

CDPD: Convención sobre los Derechos de las Personas con Discapacidad.

CEDH: Convenio Europeo de Derechos Humanos.

CHAID: *chi-square automatic interaction detection* (detección automática de la interacción de chi cuadrado).

FIM: *federal information management* (traducción estandarizada del lenguaje legal utilizado en un servicio público en Alemania).

IA: inteligencia artificial.

LLM: *large language models* (modelos de lenguaje grandes).

ML: *machine learning* (aprendizaje automático).

NLP: *natural language processing* (procesamiento de lenguaje natural).

RAG: *retrieval-augmented generation* (generación aumentada por recuperación).

RGPD: Reglamento General de Protección de Datos (Unión Europea).

RPA: *robotic process automation* (automatización robótica de procesos).

XAI: *explainable artificial intelligence* (inteligencia artificial explicable).

Fuente: elaboración propia

