Exploración, análisis y prospección en la aplicación de la IA en los procesos de evaluación de las políticas públicas. Informe de resultados.

Abril 2025

besaldi
Enplegu eta inklusio politiken ebaluazio organoa
Órgano de evaluación de las políticas de empleo e inclusión

EUSKO JAURLARITZA
GOBIERNO VASCO
EKONOMIA, LAN ETA ENPLEGU SAILA
DEPARTAMENTO DE ECONOMÍA, TRABAJO Y EMPLEO

besaldi
Enplegu eta inklusio politiken ebaluazio organoa
Órgano de evaluación de las políticas de empleo e inclusión

EUSKO JAURLARITZA
GOBIERNO VASCO
EKONOMIA, LAN ETA ENPLEGU SAILA
DEPARTAMENTO DE ECONOMÍA, TRABAJO Y EMPLEO

Exploración, análisis y prospección en la aplicación de la IA en los procesos de evaluación de las políticas públicas

Informe de resultados

[abril de 2025]

Elaborado con la asistencia técnica de Fresno the right link

Tabla de contenido

0. INTRODUCCIÓN 5

1. OBJETO DEL ESTUDIO 7

1.1 APROXIMACIÓN CONCEPTUAL 7
1.2 MARCO INSTITUCIONAL 11
1.3 ÁMBITO TEMÁTICO 12

2. MARCO METODOLÓGICO 16

3.1 PREGUNTAS DE INVESTIGACIÓN 16
3.2 REVISIÓN EXPLORATORIA 16
3.3 CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN 17
3.4 PROCEDIMIENTO DE SELECCIÓN 18
3.5 ARTÍCULOS SELECCIONADOS 19

3. PRINCIPALES HALLAZGOS 21

3.1 APLICACIÓN DE IA EN LA IMPLEMENTACIÓN DE POLÍTICAS 22
3.2 APLICACIONES DE IA EN LA EVALUACIÓN DE POLÍTICAS 28
3.3 IMPLICACIONES PARA LAS ORGANIZACIONES EVALUADORAS 35

4. TENDENCIAS Y RETOS 39

4.1 PRINCIPALES VENTAJAS 39
4.2 RETOS Y DESAFÍOS 40
4.3 REFLEXIONES DESDE LA PRÁCTICA EVALUATIVA 42

5. LIMITACIONES 45

5.1. LIMITACIONES DE LA LITERATURA ANALIZADA 45
5.2. LIMITACIONES DEL ESTUDIO 46

6. BIBLIOGRAFÍA Y REFERENCIAS 49

6.1. BIBLIOGRAFÍA 49
6.2. REFERENCIAS INCLUÍDAS EN LA MUESTRA DE LA REVISIÓN 50

7. ANEXOS METODOLÓGICOS 54

7.1. ANEXO I. ALGORITMOS BOOLEANOS DE BÚSQUEDA 54
7.2. ANEXO II. RESUMEN DE HALLAZGOS EN EL ÁMBITO DE LA IMPLEMENTACIÓN 58
7.3. ANEXO III. RESUMEN DE HALLAZGOS EN EL ÁMBITO DE LA EVALUACIÓN 62

Ilustraciones

Ilustración 1. Sistema de IA en etapa de construcción y de uso 8

Ilustración 2. Datos y tecnologías en los sistemas modernos. 10

Ilustración 3 El ciclo de las políticas..... 12

Ilustración 4 - Proceso de selección.....20

Ilustración 5. Evolución de la actitud hacia la aplicación de IA en las políticas públicas..... 21

Ilustración 6. Tecnologías IA adoptadas por región según la muestra revisada 29

Ilustración 7. Ejemplo de IA generativa para la visualización de datos.....35

Ilustración 8. Flujos de trabajo Humano-IA..... 36

Ilustración 9. Marco de competencias y riesgo de automatización..... 38

Ilustración 10. Tensiones en la aplicación de la IA..... 39

Ilustración 11 Principales tendencias..... 40

Ilustración 12. Proceso de integración de tecnologías y barreras 42

Ilustración 13 - Propuestas de mejora y lecciones aprendidas..... 48

Tablas

Tabla 1. Criterios de inclusión y exclusión..... 17

Tabla 2. Criterio de selección (relevancia).....18

Tabla 3 Número de artículos mencionando casos de implementación, por país.....22

Tabla 4 - Resumen de hallazgos en la primera revisión de implementación..... 58

Tabla 5 Resumen de hallazgos en casos de uso para evaluación (casos de uso). 62

0. Introducción

El presente informe presenta la síntesis del **estudio exploratorio sobre buenas prácticas, innovaciones y transformaciones en la aplicación de la Inteligencia Artificial en el ciclo de la política pública**, encargado por **Besaldi**, órgano de evaluación de las políticas de empleo e inclusión del País Vasco.

Besaldi es una entidad independiente adscrita al Gobierno Vasco, creada en 2024 que tiene la misión de contribuir sistemáticamente al conocimiento y perfeccionamiento de los diagnósticos, decisiones y actuaciones gubernamentales en la garantía y protección del derecho al empleo y a la inclusión social.

Entre las competencias principales de Besaldi destaca el seguimiento y evaluación de las políticas de empleo e inclusión, el desarrollo de metodologías y estándares de calidad, la promoción de investigación sobre necesidades e intervenciones, y la difusión de información sobre el funcionamiento y alcance de las políticas públicas.

La necesidad de este estudio surge específicamente en el marco del proyecto estratégico 2025-2029 de Besaldi, en línea con la dinámica transversal relacionada con la *Interacción con inteligencias artificiales*.

Para responder al propósito de la indagación y en coherencia con las tareas previstas en el encargo, el estudio se ha realizado a través de una **revisión bibliográfica exploratoria o scoping review**, y se ha centrado en **experiencias de uso documentadas** en el ciclo de la política pública, específicamente en las **fases de implementación y evaluación**. La revisión se ha centrado en la literatura académica y gris.

El informe describe los principales resultados de la revisión, y se articula en torno a seis apartados.

- El **primer capítulo** describe el objeto del estudio, delimitando la conceptualización funcional e institucional de la Inteligencia Artificial, las políticas públicas de empleo e inclusión, el ciclo de la política y las fases del proceso evaluativo.
- El **segundo apartado** proporciona el marco metodológico, explicando la revisión exploratoria realizada, las fuentes consultadas, los criterios de inclusión/exclusión aplicados y el proceso sistemático de selección de artículos mediante el protocolo de Scoping Review ajustado a los criterios PRISMA.
- El **tercer capítulo** ilustra los principales hallazgos, centrados en las fases de implementación y de evaluación de las políticas de empleo e inclusión. En cada fase, se han identificado las aplicaciones en casos de uso agrupados, que permiten apreciar los principales patrones y tendencias.
- A continuación, en la **sección cuatro**, se discuten las principales tendencias observadas en la literatura, con especial hincapié en las oportunidades, riesgos y desafíos; y para finalizar, en los capítulos **cinco** y **seis** se ilustran, respectivamente, las principales limitaciones del estudio y las referencias bibliográficas. Al final del documento se pueden encontrar los **Anexos Metodológicos**.

El informe se complementa con una recopilación de 22 fichas ilustrativas de casos, incluidas en el documento dedicado "Casos de uso".

A pesar de la metodología sistemática empleada en este estudio, es importante reconocer sus limitaciones inherentes: experiencias innovadoras recientes, iniciativas experimentales sin documentación formal y proyectos descritos con terminología específica pueden haber quedado fuera del alcance de la revisión. No obstante, el trabajo ofrece un mapeo representativo que permite identificar tendencias significativas y elementos esenciales en la aplicación de IA en las políticas de empleo e inclusión social.

1. Objeto del estudio

1.1 Aproximación conceptual

Para analizar el papel de la Inteligencia Artificial (IA) en el ciclo de las políticas públicas partimos de una aproximación conceptual a la IA, cuya definición inicial se ha consolidado a lo largo del estudio en base a la revisión bibliográfica realizada.

1.1.1 Definición

Si bien no existe una definición formal y universalmente aceptada de Inteligencia Artificial (IA), los principales marcos institucionales y regulatorios, como se verá, coinciden en que se trata de *sistemas capaces de analizar datos, inferir patrones y generar resultados o tomar decisiones con cierto grado de autonomía para alcanzar objetivos específicos.*

La **OECD** ha definido un sistema de IA como un sistema basado en máquinas que, para objetivos explícitos o implícitos, infiere, a partir de los datos que recibe, cómo generar resultados tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Los diferentes sistemas de IA varían en sus niveles de autonomía y capacidad de adaptación después de su implementación[1].

Por otro lado, **la Comisión Europea** define la Inteligencia Artificial como sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción –con cierto grado de autonomía– con el fin de alcanzar objetivos específicos. Los sistemas basados en la IA pueden consistir simplemente en un programa informático o estar incorporados en dispositivos de hardware[2].

La **estrategia española de Inteligencia Artificial**[3], por su parte, entiende la Inteligencia Artificial como sistemas que pueden resolver problemas complejos, aplicando técnicas avanzadas de programación, sin necesidad de establecer de antemano los pasos o instrucciones a seguir, tal y como sucede en los programas informáticos convencionales. Los sistemas de IA reciben como entrada múltiples datos (casos o ejemplos) y, a partir de ellos, ajustan sus modelos internos hasta encontrar la solución a un problema.

A **nivel funcional**, por tanto, se puede entender la IA como un **programa informático** capaz de reaccionar como lo haría una persona inteligente, entendiendo como programa informático un conjunto de instrucciones escritas por un programador, que son interpretadas por un ordenador para realizar diversas tareas. Según el **estándar internacional ISO/IEC 22989**[4], la IA implica el desarrollo de mecanismos y aplicaciones que emulan aspectos de la inteligencia humana y que, a través de una implementación ingenierizada, genera

[1] OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449.
[2] Comisión Europea. (2018). *Inteligencia artificial para Europa* (COM(2018) 237 final). Recuperado del siguiente enlace.
[3] https://portal.mineco.gob.es/es-es/digitalizacionIA/Documents/Estrategia_IA_2024.pdf.
[4] International Organization for Standardization & International Electrotechnical Commission. (2022). Information technology — Artificial intelligence — Artificial intelligence concepts and terminology (ISO/IEC Standard No. 22989:2022(E)).

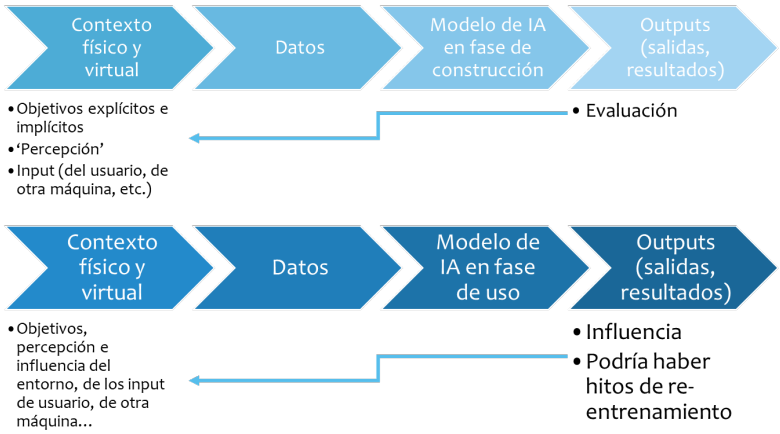
resultados como contenido, pronósticos, recomendaciones o decisiones orientados a cumplir objetivos específicos definidos por humanos.

En este sentido, **a diferencia de la programación clásica, la IA no sigue reglas predefinidas** por los programadores, sino que desarrolla sus propios patrones de reconocimiento a partir de datos etiquetados. En la programación clásica, un desarrollador debe especificar explícitamente cada paso computacional requerido para resolver un problema. La lógica es predeterminada y sigue un flujo predecible de instrucciones. En contraste, los sistemas de IA, especialmente aquellos basados en aprendizaje automático, derivan su lógica parcial o totalmente de los datos utilizados para su entrenamiento. Los cálculos o pasos necesarios para resolver un problema no están determinados de antemano, sino que emergen del proceso de aprendizaje.

En definitiva, la Inteligencia Artificial se caracteriza por procesar la información a través de datos de entrenamiento sin reglas necesariamente predefinidas, emulando así capacidades de la inteligencia humana para la toma de decisiones.

Como se puede apreciar en la **Ilustración 1**, de hecho, un sistema de IA recibe datos del entorno y los procesa sin reglas predefinidas: en lugar de tener instrucciones codificadas por programadores que digan "si ocurre X, entonces haz Y", el sistema desarrolla sus propios modelos de análisis y toma de decisiones basándose en los datos que recibe. Este proceso se hace evidente en la parte derecha de la ilustración, donde se observa cómo el sistema aplica estos patrones aprendidos para procesar nueva información. La flecha de "influencia" hacia el entorno muestra el resultado de estas decisiones autónomas basadas en patrones aprendidos, no en reglas programadas.

Ilustración 1. Sistema de IA en etapa de construcción y de uso



Fuente: elaboración propia adaptado desde OECD AI Principles overview.

En la IA las reglas que el sistema desarrolla durante este aprendizaje suelen ser opacas, aunque demuestren ser efectivas cuando se aplican a nuevos casos no vistos durante el entrenamiento. Este **proceso de aprendizaje automático** representa un cambio de paradigma respecto a la programación secuencial tradicional, pasando de un enfoque basado

en reglas explícitas a uno basado en el reconocimiento de patrones mediante el procesamiento de grandes cantidades de datos etiquetados.

Otra distinción significativa es la **capacidad evolutiva de los sistemas de IA**. Mientras que el software tradicional requiere ser reescrito para incorporar mejoras o adaptaciones, los sistemas de IA pueden perfeccionarse con el tiempo sin necesidad de modificar su código base. Esto se logra mediante el reentrenamiento con nuevos datos y la optimización de parámetros del modelo, permitiendo una adaptabilidad que emula ciertos aspectos del aprendizaje natural.

La capacidad de aprendizaje y adaptación también implica desafíos particulares en términos de **verificación, validación y explicabilidad**, que el estándar ISO/IEC 22989 reconoce como elementos esenciales en el ciclo de vida de los sistemas de Inteligencia Artificial. En modelos complejos, especialmente los que aprenden a partir de grandes volúmenes de datos, estos procesos son menos directos que en el software convencional y requieren enfoques especializados para garantizar que el sistema sea confiable, transparente y seguro.

1.1.2 Evolución y clasificación

Como se ha mencionado anteriormente, l**a IA puede entenderse como un conjunto de herramientas y programas informáticos que intentan imitar algunas capacidades de la mente humana.** Esta disciplina ha evolucionado de forma no lineal a lo largo de las últimas décadas, con periodos de gran entusiasmo alternados con etapas de menor avance (conocidos como "inviernos de la IA").

En sus inicios, predominó la denominada **IA Simbólica**, que se basa en la manipulación de símbolos y reglas lógicas predefinidas para la toma de decisiones. Este enfoque fue pionero en el desarrollo de programas como ELIZA (1964), que simulaba una conversación terapéutica mediante patrones de coincidencia simples y respuestas preprogramadas. Aunque ELIZA creaba la ilusión de comprensión, no contaba con un verdadero conocimiento sobre psicoterapia. La IA Simbólica presenta limitaciones importantes, ya que depende de una codificación manual del conocimiento y carece de capacidad de aprendizaje autónomo. En la actualidad, algunos sistemas de evaluación crediticia siguen empleando principios de este enfoque.

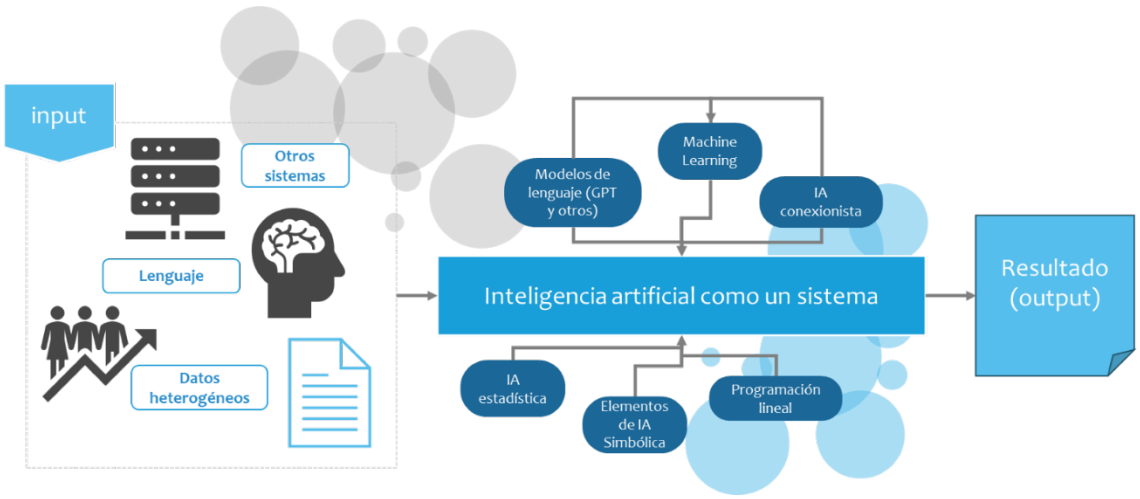
Paralelamente, y con mayor auge desde los años 80, se desarrolló la **IA Conexionista**, basada en redes neuronales artificiales que imitan, en cierto grado, el funcionamiento del cerebro humano. A diferencia del enfoque simbólico, estos modelos pueden aprender de manera autónoma a partir de grandes volúmenes de datos, lo que los hace particularmente eficaces en tareas como el reconocimiento de imágenes y el procesamiento del lenguaje natural. El resurgimiento de este enfoque en la última década, gracias al aumento de la capacidad computacional y la disponibilidad de datos masivos, ha impulsado enormemente el campo de la IA.

Es importante destacar que ambos enfoques — el simbólico y conexionista— han contribuido al desarrollo del **aprendizaje automático (*Machine Learning*)**, un campo fundamental de la IA que engloba diversos métodos algorítmicos. El aprendizaje automático puede entenderse como un conjunto de técnicas que permiten a los ordenadores "aprender"

a partir de ejemplos y mejorar con la experiencia, sin necesidad de ser programados explícitamente para *cada* tarea específica. Mientras que algunas técnicas de aprendizaje automático, como las redes neuronales profundas, son claramente conexionistas, otras como los árboles de decisión o los sistemas basados en reglas tienen elementos más cercanos al enfoque simbólico.

El aprendizaje automático incluye un amplio espectro de técnicas como regresiones, métodos bayesianos, máquinas de vectores soporte y algoritmos de agrupamiento, entre otros. La característica común de todos estos métodos es su capacidad para identificar **patrones** en los datos y utilizarlos para hacer predicciones o, incluso, ayudar en la toma de decisiones ante nuevos datos. En el contexto de las políticas públicas, estas herramientas ofrecen aplicaciones muy valiosas: desde la estratificación de poblaciones para focalizar ayudas sociales, hasta la detección de fraudes en sistemas de prestaciones o el diseño de intervenciones basadas en evidencia para programas de empleo.

Ilustración 2. Datos y tecnologías en los sistemas modernos.



Fuente: Elaboración propia.

Actualmente, la evolución de la IA apunta hacia **enfoques híbridos**, que combinan la interpretabilidad de la IA Simbólica con la capacidad de aprendizaje de la IA Conexionista. Estos sistemas buscan abordar uno de los principales desafíos de los modelos más avanzados: su naturaleza de "caja negra", que dificulta comprender cómo llegan a determinadas conclusiones. La **explicabilidad** o **interpretabilidad** de los modelos resulta crucial en aplicaciones donde la transparencia y la rendición de cuentas son fundamentales, como es el caso de la evaluación de políticas públicas que afectan directamente a la ciudadanía.

Una **clasificación** relevante para este estudio distingue la IA **según su capacidad**. En la actualidad, todos los sistemas disponibles corresponden a la **IA Estrecha o Narrow**, diseñada para realizar tareas específicas en dominios limitados, como reconocimiento facial, traducción automática o análisis predictivo, aunque la atención recaiga, sobre todo, en aquellos modelos estrechos para la generación de contenidos (siendo el caso más conocido hoy el de los modelos GPT, o *Generative Pre-trained Transformer*). La **IA General**, con habilidades equiparables a las humanas en diversos dominios y capacidad de transferir

conocimiento entre ellos, permanece como un objetivo técnico potencialmente viable pero aún no alcanzado. Por su parte, la **Superinteligencia**, que teóricamente superaría las capacidades cognitivas humanas en prácticamente todos los aspectos, sigue siendo una hipótesis controvertida más que una realidad cercana.

1.2 Marco institucional

La regulación de la Inteligencia Artificial (IA) se desarrolla en distintos niveles administrativos, cada uno con capacidades vinculantes y ámbitos de aplicación específicos. En este contexto, la **Organización para la Cooperación y el Desarrollo Económico (OCDE)** estableció en 2019 los primeros **principios intergubernamentales sobre IA**[5]. Aunque no tienen fuerza legal vinculante, han servido como una referencia clave en la formulación de regulaciones posteriores. Estos principios, centrados en el crecimiento inclusivo, la equidad, la transparencia, la seguridad y la responsabilidad, han influido en marcos regulatorios y estrategias nacionales e internacionales. En 2024, la OCDE revisó sus principios[6], reforzando la supervisión, la transparencia y la sostenibilidad, con el objetivo de garantizar un desarrollo responsable y alineado con los derechos fundamentales. Se introdujeron también directrices para la IA generativa y el uso de datos sintéticos, además de nuevos criterios para evaluar riesgos.

A **nivel europeo**, la Unión Europea ha adoptado un enfoque normativo integral con la aprobación en 2023 del **AI Act**[7], el primer reglamento vinculante sobre IA en la UE. Como reglamento, el AI Act es de aplicación directa en los Estados miembros sin necesidad de transposición, estableciendo un marco basado en la clasificación de riesgos, con obligaciones específicas y mecanismos de supervisión. Su implementación es progresiva, con la entrada en vigor de sus disposiciones principales en 2025, aunque algunas obligaciones específicas se aplicarán en plazos diferenciados. Además, la UE ha complementado este marco normativo con estrategias de IA enfocadas en la inversión en innovación, el fomento de la investigación y el establecimiento de estándares éticos.

En el **ámbito nacional**, España presentó en 2020 su **Estrategia Nacional de Inteligencia Artificial (ENIA)**[8], un documento estratégico que orienta el desarrollo de la IA en el país. Si bien la ENIA no es una norma con carácter vinculante, sí establece un marco de referencia para las políticas de inversión, desarrollo tecnológico y regulación futura. La estrategia pone especial énfasis en la capacitación digital, el fortalecimiento del ecosistema de innovación y la generación de entornos de confianza. En términos normativos, España ha desarrollado iniciativas para adaptar su legislación a los requerimientos del AI Act, incluyendo la creación de organismos de supervisión y la regulación de aspectos específicos como el uso de IA en el sector público.

A **nivel autonómico**, la Comunidad Autónoma del País Vasco ha desarrollado el **Plan Vasco de Inteligencia Artificial (BAHI 2024)**[9], que adapta las estrategias de IA a su realidad territorial y

[5] https://oecd.ai/en/ai-principles.
[6] Ibidem.
[7] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689.
[8] https://portal.mineco.gob.es/es-es/digitalizacionIA/Documents/Estrategia_IA_2024.pdf.
[9] https://estrategia.baic.eus/.

se centra en la implementación práctica dentro de sectores estratégicos y en la colaboración público-privada. El BAHI 2024 complementa el marco normativo con programas de inversión en I+D, formación especializada y creación de infraestructuras tecnológicas que permitan a las empresas y administraciones públicas integrar soluciones de IA de manera eficiente y ética.

Mientras que el marco normativo establece obligaciones vinculantes, las estrategias nacionales y autonómica complementan esta regulación a través de políticas públicas que fomentan la innovación, la adaptación territorial y la armonización con los principios éticos internacionales. Este modelo regulador y estratégico busca garantizar un desarrollo de la IA que combine seguridad jurídica, flexibilidad y respeto a los derechos fundamentales, promoviendo un equilibrio entre regulación, innovación y capacidad de adaptación a contextos específicos.

1.3 Ámbito temático

El presente estudio explora la potencial aplicación de la IA en el ciclo de las **políticas públicas** de **empleo e inclusión**, con especial referencia a los procesos de **formulación** y **evaluación** de políticas públicas.

Ilustración 3 El ciclo de las políticas

Fuente: Elaboración propia.

Como se verá posteriormente, los principales hallazgos y casos de uso de la IA en el ámbito de las políticas de empleo e inclusión social, se encuentran en la fase de **implementación** incluyendo la provisión de servicios, dentro del ciclo de las políticas, y en el marco de la **operacionalización y síntesis** dentro del proceso de evaluación.

1.3.1 *Las políticas de empleo e inclusión social*

El estudio se centra en la intersección entre la Inteligencia Artificial y dos ámbitos fundamentales de política pública en el contexto vasco: las políticas de empleo y las políticas de inclusión social. Estos ámbitos representan pilares esenciales del bienestar ciudadano que enfrentan transformaciones sustanciales en el actual contexto de cambio tecnológico, económico y social.

Las **políticas de empleo** constituyen el conjunto de intervenciones públicas destinadas a mejorar el funcionamiento del mercado laboral, promover la empleabilidad de las personas y facilitar su inserción y permanencia en un trabajo digno. En el contexto vasco, estas políticas se articulan fundamentalmente a través de la Red Vasca de Empleo, con Lanbide (Servicio Vasco de Empleo) como actor central. Este ámbito adquiere especial relevancia en un escenario caracterizado por la transformación digital de la economía, la automatización creciente y la necesidad de adaptación continua de competencias profesionales. Las políticas de empleo deben responder al desafío de garantizar que el trabajo humano mantenga su valor como actividad productiva, como fuente de ingresos y como elemento de identidad personal y cohesión social, en un contexto donde las transformaciones tecnológicas, ecológicas y sociales reconfiguran profundamente la naturaleza y distribución del trabajo.

Las **políticas de inclusión social** comprenden el conjunto de medidas orientadas a prevenir y abordar situaciones de vulnerabilidad y exclusión, promoviendo la plena participación de todas las personas en la vida económica, social y cultural. En Euskadi, estas políticas encuentran su expresión institucional en el Sistema Vasco de Garantía de Ingresos y para la Inclusión, y operan de manera transversal a través de diversos ámbitos sectoriales como empleo, garantía de ingresos, educación, salud, vivienda y servicios sociales. Su objetivo principal se centra en la protección frente a la pobreza y la lucha contra la exclusión social en sus múltiples dimensiones. Estas políticas afrontan el reto de adaptarse a un horizonte marcado por transformaciones significativas en la matriz productiva, en el modelo de cuidados y en la configuración del sector público, buscando garantizar la inclusión efectiva de todos los ciudadanos en un contexto de creciente complejidad y desigualdad 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.

1.3.2 *El ciclo de las políticas públicas*

El **ciclo de políticas públicas** representa un marco conceptual que estructura el proceso de diseño, implementación y evaluación de intervenciones gubernamentales. Este ciclo tradicionalmente se compone de **cinco fases interconectadas** que operan de manera iterativa, configurando un proceso dinámico y continuo de mejora en la acción gubernamental (Subirats, 2008).

Fase 1. Identificación del problema. En esta etapa inicial, los actores públicos identifican una situación problemática o una necesidad que requiere de intervención pública. Se define el problema público, se analizan sus causas y se determina su magnitud e impacto social. Esta

fase implica la recopilación y análisis de datos, la consulta con grupos afectados y la colocación del problema en la agenda pública. El resultado es una clara delimitación del problema que justifica la intervención pública.

Fase 2. Formulación de alternativas. Durante esta etapa se desarrollan diferentes opciones de políticas para abordar el problema identificado. Se consideran diversas estrategias, instrumentos y mecanismos de intervención. Este proceso incluye análisis técnicos, valoración de restricciones legales y presupuestarias, y consideración de experiencias previas o comparadas. El objetivo es contar con un catálogo de posibles soluciones técnicamente viables y políticamente factibles.

Fase 3. Toma de decisiones. Esta fase implica la selección de una alternativa específica entre las opciones disponibles. Conlleva un proceso de deliberación que considera factores técnicos, políticos, económicos y sociales. Se evalúan los costes y beneficios de cada alternativa, se realizan procesos de negociación entre actores con diferentes intereses y se legitima la opción elegida. El resultado es la formalización de la política a través de instrumentos normativos apropiados.

Fase 4. Implementación. Esta etapa comprende la puesta en marcha de la política seleccionada. Incluye la asignación de recursos, la definición de procedimientos operativos y la coordinación entre diferentes actores y niveles administrativos. Durante esta fase se materializa la política mediante actividades concretas, se gestiona la red de implementadores y se realiza un seguimiento de la ejecución. Esta fase determina el paso de la política como diseño a la política como realidad operativa.

Fase 5. Evaluación. En esta fase final se valora el impacto y los resultados de la política implementada. Se analiza la eficacia, eficiencia y pertinencia de la intervención, generando información para futuros ciclos de políticas. Esta evaluación puede ser formativa (durante la implementación) o sumativa (una vez concluida), y puede centrarse en diferentes aspectos como el diseño, los procesos o los resultados. Los hallazgos de esta fase alimentan un nuevo ciclo de políticas, completando así la naturaleza iterativa del proceso.

1.3.3 El proceso evaluativo

El estudio se centra en el ciclo de la política en general, especialmente en los procesos relacionados con la fase de evaluación. A su vez, la evaluación de políticas públicas se articula en un proceso sistemático y estructurado que comprende seis fases interrelacionadas, de acuerdo a la teoría y práctica evaluativa (Better Evaluation, 2014; European Commission, 2024; Ligero Lasa, 2015).

Fase 1. Encargo y gestión de la evaluación. El proceso evaluativo comienza con la formulación del encargo por parte de una institución pública, organismo internacional o entidad privada. En esta etapa se definen los propósitos generales de la evaluación, estableciendo qué se quiere conseguir con el estudio y para qué se utilizarán sus resultados. Se identifican las partes interesadas (stakeholders) y se negocian las expectativas, alcance, recursos disponibles, plazos y limitaciones. Esta fase es crucial pues determina el enfoque general y establece la relación contractual entre el equipo evaluador y la entidad que solicita la evaluación.

Fase 2. Definición del objeto de la evaluación. Esta fase implica la delimitación precisa de qué se va a evaluar, contextualizando la política o programa dentro de su marco institucional, normativo y socioeconómico. Se elabora una "teoría del cambio" o modelo lógico que explicita cómo se espera que la intervención produzca los resultados deseados, identificando insumos, actividades, productos, resultados e impactos previstos. Se definen los criterios evaluativos y las preguntas clave que guiarán la investigación. Es fundamental comprender la historia de la política, sus objetivos explícitos e implícitos, y el contexto político en que se desarrolla.

Fase 3. Elección del enfoque evaluativo. En esta fase se adopta una posición paradigmática fundamental que orientará todo el proceso evaluativo. Se trata de una decisión ontológica y epistemológica sobre cómo concebir la naturaleza de la evaluación, qué se considera conocimiento válido y cuál es su finalidad última. Históricamente, los enfoques evaluativos se han desarrollado alrededor de tres dimensiones principales: el uso que se dará a la evaluación, los valores que guían el juicio evaluativo, y los métodos empleados para generar evidencia confiable. Esta elección paradigmática determina cuestiones cruciales como el rol del evaluador, la relación con los stakeholders y los criterios de calidad evaluativa.

Fase 4. Operacionalización. Esta fase convierte los planteamientos conceptuales en acciones prácticas y sistemáticas. Inicia con la elaboración de la matriz de evaluación, instrumento central que vincula preguntas, criterios, indicadores y fuentes de información, estableciendo la estructura lógica del estudio. A continuación, se diseña el trabajo de campo, definiendo la estrategia de acercamiento al territorio y a los actores: se elaboran los instrumentos de recolección específicos, se determina el muestreo, se establecen los protocolos éticos y se calendarizan las actividades. Finalmente, se ejecuta la recogida de datos mediante técnicas cuantitativas y/o cualitativas, adaptándose a las condiciones del contexto.

Fase 5. Síntesis. En esta fase se integran, analizan e interpretan los datos recopilados para generar hallazgos significativos que respondan a las preguntas de evaluación. Implica la triangulación de fuentes y métodos para reforzar la validez de las conclusiones, así como el análisis crítico de las evidencias a la luz de la teoría del cambio y los criterios evaluativos establecidos. Se identifican patrones, tendencias, relaciones causales y factores contextuales que explican los resultados observados. La síntesis no es una mera descripción de hallazgos, sino un proceso analítico que transforma datos en conocimiento accionable, culminando con la formulación de conclusiones sólidas.

Fase 6. Informe y uso. La sexta y última fase se centra en el informe y apoyo al uso de la evaluación, desarrollando y presentando los hallazgos de manera que sean útiles para los diferentes actores implicados. Esta fase es crucial para asegurar que la evaluación cumpla su propósito y genere aprendizajes significativos. Incluye la elaboración de informes adaptados a diferentes audiencias, la presentación de resultados de forma accesible y comprensible, y el acompañamiento en la utilización de los hallazgos para la mejora de la intervención.

2. Marco metodológico

3.1 Preguntas de investigación

El estudio se ha vertebrado entorno a las siguientes **preguntas de investigación**.

- ¿Cómo se está aplicando actualmente la Inteligencia Artificial en los procesos de formulación y evaluación de políticas públicas de empleo e inclusión social en el ámbito europeo y sistemas administrativos comparables/ similares al español?
- ¿Qué aplicaciones transformadoras podrían optimizar estos procesos, con especial atención a las aplicaciones potencialmente transferibles a órganos de evaluación de políticas públicas como Besaldi?

Estas cuestiones plantean un **estudio exploratorio** que pretende identificar, analizar y sistematizar tanto las aplicaciones existentes como las potencialidades emergentes de la IA en el ciclo de políticas públicas, con especial atención a los ámbitos de empleo e inclusión social.

El carácter exploratorio de la investigación responde a la naturaleza incipiente y en rápida evolución de la aplicación de estas tecnologías en la administración pública, donde coexisten experiencias pioneras con amplios espacios de desarrollo aún no explorados. Asimismo, el componente prospectivo permitirá anticipar tendencias futuras y orientar estratégicamente la incorporación de estas tecnologías.

Para abordar esta pregunta de forma sistemática y rigurosa, se ha realizado una **scoping review o revisión exploratoria** ampliada, que integra tanto el análisis del estado actual como la dimensión prospectiva.

3.2 Revisión exploratoria

La revisión **bibliográfica exploratoria o scoping review** es un tipo de revisión de la literatura usado para explorar la evidencia disponible sobre un tema. Su objetivo principal es *mapear* conceptos, identificar los recursos y la investigación realizada en diferentes contextos (sean estos académicos o no) para detectar oportunidades en la investigación actual y plantear nuevas vías de análisis y de puesta en práctica. La revisión se ha desarrollado siguiendo el marco metodológico de Arksey y O'Malley (2005) con las mejoras propuestas por Levac et al. (2010), adaptado para incorporar la dimensión prospectiva.

La scoping review ha abarcado el **estado actual** de la IA en materia de política social y, especialmente, en lo referido a empleo e inclusión social, con especial énfasis en las rentas mínimas garantizadas, ingreso mínimo vital (de aquí en adelante, IMV), Renta de Garantía de Ingresos (de aquí en adelante, RGI), figuras equiparables y términos sinónimos; y las principales **tendencias emergentes y futuras aplicaciones**.

3.3 Criterios de inclusión y exclusión

Una vez definida la pregunta de investigación que orienta las decisiones metodológicas de la revisión, se han definido los **criterios de inclusión y exclusión** para delimitar el corpus documental relevante. Dichos criterios establecen parámetros claros sobre qué tipos de documentos serán analizados, considerando aspectos como su naturaleza académica o institucional, la metodología empleada, el marco temporal, las fuentes de búsqueda y los idiomas aceptados.

Tabla 1. Criterios de inclusión y exclusión

Criterios de inclusión	Criterios de exclusión
Tipo de publicación: se pueden incluir artículos e informes de las siguientes fuentes: (i) publicaciones académicas (artículos y monografías): (ii) cualesquiera otros estudios revisados por pares, incluyendo tesis doctorales; (iii) literatura gris gubernamental y del tercer sector siempre y cuando se reflejen y relaten casos de uso concretos y transferibles; documentos institucionales, planes estratégicos, diagnósticos y mapeo de actores.	Se excluyen artículos periodísticos, artículos de blog, y artículos breves, por su falta de fundamentación empírica. Con el fin de poder acotar el objeto de la investigación, se excluyen, también, informes de política educativa, salud y salud pública, infraestructuras, y otros artículos y monografías o informes referentes a ámbitos de negocio (p.e., marketing, finanzas, etc.). Documentos anteriores a 2019.
Metodología: son admisibles estudios cuantitativos, cualitativos, mixtos y teóricos, aun con ciertas reservas con respecto a estos últimos considerando los objetivos de la Scoping review (mapear buenas prácticas y casos de uso, que podrían aparecer mencionados).	
Tiempo: se incluyen fuentes desde 2019 hasta la actualidad, debido al avance de la técnica y, también, a su adopción y generalización o adopción institucional. La elección de fechas se justifica también de acuerdo con hitos de importancia, tales como los documentos de consenso de ALTAI en cuanto a ética de la IA, legislación europea y nacional o hitos tecnológicos.	
Fuentes: se realizan búsquedas en bases de datos académicas y generalistas, incluyendo Google Académico (Scholar), Google, y revisando WOS y SCOPUS.	
Idiomas: español e inglés.	

En cuanto al **ámbito temático**, se han revisado documentos que abordan, de forma explícita, aplicaciones de la IA en políticas sociales, en evaluación de dichas políticas y su efecto y/o impactos, casos de uso en programas sociales – con especial interés en aquellos de IMV, RGI y figuras similares -, e implementaciones, aun temporales, en servicios sociales de empleo e inclusión.

La búsqueda sistemática se ha realizado en los motores y fuentes anteriormente detallados con **algoritmos booleanos replicables y estructurados** que se detallan con más rigor en el Anexo I, y que se han adaptado a fuentes tanto académicas como generalistas. Esta duplicidad de fuentes es oportuna si perseguimos el objetivo de cubrir también la literatura institucional y los casos de uso que se han probado y documentado en la práctica.

La búsqueda se ha centrado en términos tales 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.*

3.4 Procedimiento de selección

El procedimiento para la selección final de las publicaciones y la extracción de la información ha comprendido varias fases secuenciales:

1. **Exploración preliminar** de los resultados de la búsqueda mediante *skimming* (lectura diagonal) de títulos y *abstract*. El *skimming* permite una valoración rápida de la potencial relevancia del documento, pero no es un criterio suficiente de inclusión, pues no asegura la pertinencia geográfica, o de alcance, e incluso temática, una vez revisado más cuidadosamente el resultado.
2. **Aplicación de los criterios de inclusión y exclusión**, i.e., tipo de publicación, aproximación metodológica, temporalidad, idiomas y alcance geográfico. Al comprender literatura gris, también había que revisar con más cautela la naturaleza del texto.
3. **Priorización.** La priorización se ha basado en el criterio de relevancia, operacionalizado en una escala categórica de 0 (relevancia mínima / no relevante) a 4 (relevancia alta) sobre los *abstract* que iban siendo refinados en las sucesivas iteraciones.
4. **Revisión detenida de la literatura.** Se ha realizado una revisión detenida sobre los artículos muy relevantes (puntuación de 3 a 4), siguiendo un enfoque ágil y riguroso al tiempo, respondiendo a los principios de: relevancia contextual, casos concretos de uso y transferibilidad y actualidad normativa/legal.

El **criterio de relevancia** ha permitido identificar aquellos estudios con mayor potencial analítico (puntuaciones 3-4) para su revisión exhaustiva, priorizando casos de implementación práctica con significativa transferibilidad en el contexto europeo. Su operacionalización se ha establecido mediante una escala categórica de cuatro niveles (0-4) que evalúa la conexión específica entre aplicaciones de IA y políticas públicas de empleo y/o inclusión social.

Tabla 2. Criterio de selección (relevancia)

Puntuación	Nivel de relevancia	Descripción
0	Relevancia mínima o no relevante	Satisface criterios formales de inclusión, pero presenta contenido tangencial o excesivamente teórico sin conexión operativa con la aplicación de IA en políticas públicas de empleo/inclusión social. La descontextualización del marco práctico-aplicativo limita su utilidad para el análisis comprehensivo.
1	Poco relevante	Aborda la implementación de IA desde perspectivas generalistas sin especificidad contextual europea o presenta aplicabilidad restringida en marcos de empleo/inclusión. Su contribución es limitada por divergencias significativas en modelos asistenciales o normativas incompatibles con el ámbito geográfico estudiado.
2	Relevante	Incorpora elementos significativos sobre aplicación de IA en políticas de empleo/inclusión con aportaciones teórico-empíricas valiosas. Su orientación predominantemente conceptual proporciona fundamentos sin profundizar en mecanismos específicos de formulación o evaluación de políticas públicas en estos sectores.

Puntuación	Nivel de relevancia	Descripción
3 a 4	**Muy relevante**	Presenta pertinencia crítica sobre implementación de IA en formulación/evaluación de políticas de empleo/inclusión social con casos específicos variables. Contribuye significativamente a comprender aplicaciones prácticas de IA en estos ámbitos dentro del contexto europeo, con alto potencial de transferibilidad metodológica e impacto.

3.5 Artículos seleccionados

Desde **SCOPUS** y **WOS** (las dos principales bases de datos académicas) aparecieron, en total, **580** referencias: concretamente, 208 y 372 respectivamente. Tras eliminar duplicados, resultó en 482 instancias. Desde **Google académico**, aparecieron 8740 resultados y finalmente se seleccionaron 34 relevantes.

De ellos, se han priorizado 30 artículos (respectivamente, 11 de bases de datos académicas y 19 de Google académico) que se han valorado como muy relevantes (nivel 3 y 4 de priorización por relevancia).

Debido al carácter general y muy amplio de las booleanas de la primera tanda de búsqueda, se decide focalizar un nuevo ciclo de búsqueda centrado en la evaluación dentro del ciclo de políticas públicas y cómo la IA y los datos masivos pueden servir a dicha tarea evaluadora. Como punto de partida, se lee y analiza una fuente de reciente publicación (Bohni Nielsen et al., 2024) y conforme a los términos usados se diseñan los algoritmos de búsqueda.

Se extraen también algunas de sus citas (16 citas extraídas del libro) y se introducen en Litmaps para buscar por bola de nieve y conveniencia las citas de las citas atendiendo a criterio temporal (posterior a 2019) y a título (data, IA, generative AI, LLM, big data, etc.). Se incluyen en Litmaps muchos artículos contextuales y teóricos necesarios para poder introducir la cuestión, pero no incluidos en la muestra analizada finalmente obteniendo 24 resultados.

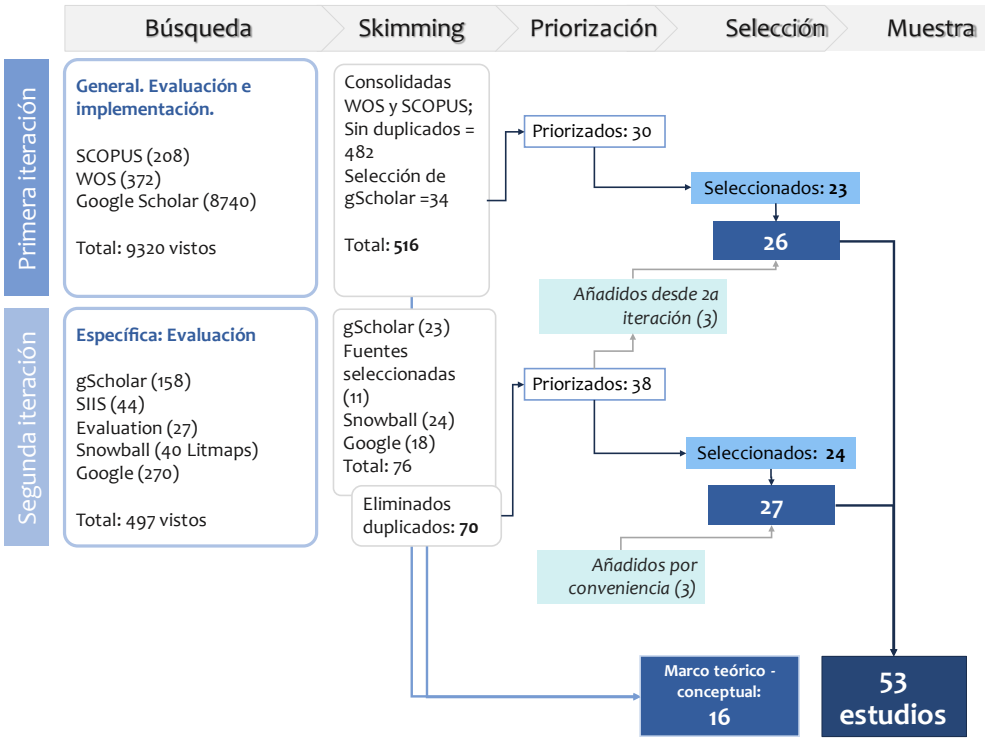
Se llevan a cabo tres búsquedas específicas: una en **Google Scholar**, con 158 resultados, y en **Google** (motor de búsqueda de propósito general) con 270 resultados; también se buscó de forma particularizada en **SIIS**, obteniendo 44 resultados, y en **Evaluation** (revista académica) obteniendo 27 resultados).

Tras eliminar duplicados y examinar títulos y abstracts quedaron 70 resultados, de los cuales fueron priorizados **38**: de ellos, uno fue inaccesible y 3 fueron movidos a los resultados de la búsqueda anterior sobre implementación. Finalmente, fueron seleccionados 24 estudios; se añadieron 3 estudios más: dos, fruto de desglosar dos capítulos (York, 2024; Ziulu et al., 2024) de una obra reciente sobre evaluación, datos masivos e inteligencia artificial (Bohni Nielsen et al., 2024) y otro que apareció incidentalmente por notificación temática en gScholar publicado a finales de Febrero de este mismo año (Babšek et al., 2025).

La selección final estaba formada, entonces, por 27 resultados, 16 de ellos específicamente focalizados en evaluación.

- Se tienen en cuenta también 4 (uno de los cuales desglosa a 8 artículos) como antecedentes, con fecha inferior al corte de la revisión de alcance actual, pero valiosos conceptualmente.
- Hay que puntualizar que no todos los artículos individuales contienen un caso de uso, o no necesariamente un caso de uso distinto.
- Se consideran a lo largo de toda la revisión 16 referencias que pertenecerían más bien a la bibliografía teórico-conceptual, y que aparecen mencionadas de forma reiterada y sistemática.

Ilustración 4 - Proceso de selección



Fuente: elaboración propia.

En total, se analizan 53 estudios, junto con el marco conceptual y la bibliografía de referencia más relevante. Las fuentes seleccionadas se han analizado temáticamente a través de la herramienta de análisis cualitativo MAXQDA (versiones 2024 y 2022), y los resultados se incluyen a continuación.

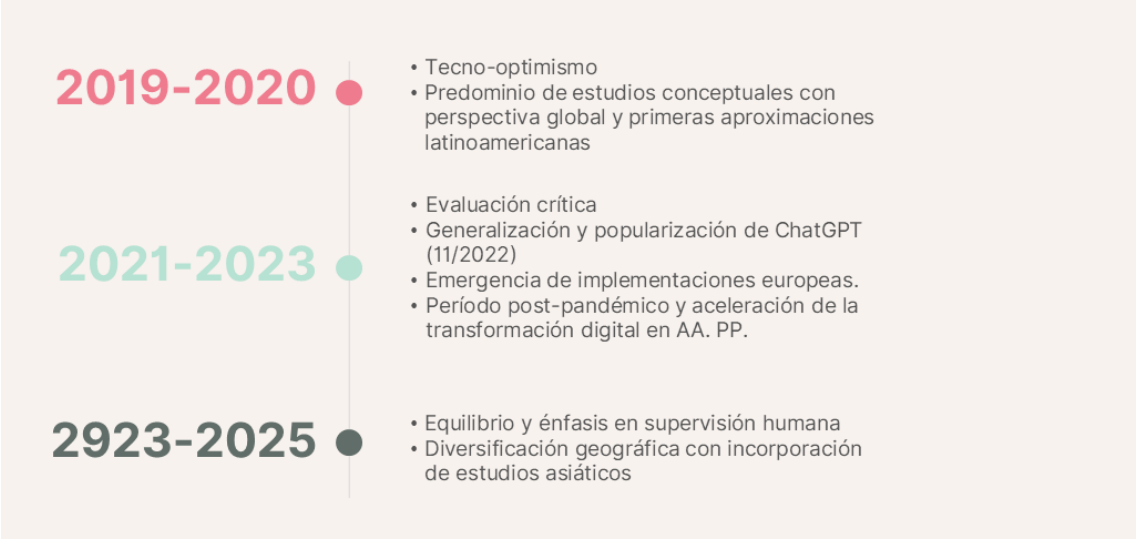
Asimismo, a lo largo del presente informe se ha utilizado de forma extensiva Inteligencia Artificial generativa; de forma específica, se han utilizado dos modelos LLM en su aplicación de uso general (chat): concretamente, Claude Sonnet 3.7 Extended y Gemini 2.5 Pro.

3. Principales hallazgos

La investigación de la aplicación de la IA en políticas públicas está en crecimiento, demostrando una creciente madurez, con diferencias importantes entre países. Los casos de uso de mayor relevancia son también aquellos de incorporación más reciente.

En este sentido, y respecto a la evolución temporal, hemos observado una evolución en la **actitud y percepción** de la IA en este ámbito, desde una postura más optimista hacia un paradigma de salvaguarda y control.

Ilustración 5. Evolución de la actitud hacia la aplicación de IA en las políticas públicas



Fuente: elaboración propia.

En cuanto a las **tecnologías** descritas en los estudios, debido al tiempo enmarcado en la búsqueda, el código dominante es el **Machine Learning o aprendizaje-máquina** junto a aseveraciones más genéricas: a pesar del auge de la IA generativa desde finales de 2022 y la irrupción de ChatGPT en la interfaz que hoy conocemos, cuando se habla de Inteligencia Artificial el Machine Learning se sigue identificando como "la" definición de IA, junto con otros términos vinculados a él: redes neurales, Deep learning, etc.

En cuanto al **ciclo de políticas**, en el ámbito de las políticas de empleo e inclusión se observa un claro predominio de aplicaciones en las fases de **implementación**, asociada en general a la **provisión de servicios** a la ciudadanía, tanto servicios sociales como servicios de empleo; y en la fase de **evaluación**. Los casos identificados en literatura revisada relativos a la formulación de políticas han sido más bien residuales (Alexander, 2022) o tratados en revisiones más generales del ciclo de política pública (Arguelles Toache, 2023; Carlizzi & Quattrone, 2023), excepto para actividades de pronóstico y diagnóstico, que se han analizado en el conjunto de la provisión de servicios. Los principales resultados se ilustran a continuación.

Finalmente, hemos dedicado un apartado específico de esta sección a las **implicaciones** que tiene la implantación de sistemas basados en IA en las entidades, tanto **a nivel organizacional** (aprendizaje), como a nivel **individual** (capacidades).

Cabe destacar que los hallazgos principales presentados en este apartado reflejan exclusivamente las iniciativas identificadas a través de la metodología sistemática empleada. Es necesario reconocer que, debido a las limitaciones inherentes de este enfoque, diversas experiencias relevantes han quedado inevitablemente fuera del alcance del estudio, incluyendo proyectos significativos implementados por distintas administraciones públicas tanto en España como internacionalmente. No obstante, los casos documentados constituyen una base representativa que permite identificar patrones claros y tendencias significativas en la aplicación de inteligencia artificial a las políticas de empleo e inclusión social.

Asimismo, en la literatura revisada se han identificado 22 casos de uso documentados, que se han sistematizado en fichas ilustrativas en el documento "Casos de uso".

3.1 Aplicación de IA en la implementación de políticas

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:

1. Eficiencia administrativa
2. Sistemas de perfilado y segmentación
3. Sistemas de detección y respuesta temprana
4. Participación ciudadana

Estas aplicaciones representan un espectro evolutivo desde la automatización administrativa básica hasta sistemas más complejos de apoyo a la decisión y detección temprana de necesidades sociales.

La **eficiencia administrativa mediante automatización** constituye el marco predominante de implementación actual. Se observa una progresión desde funciones administrativas básicas (Automatización Robótica de Procesos o RPA, 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

Tabla 3 Número de artículos mencionando casos de implementación, por país

Países (EU) con casos de implementación	*Nº de artículos revisados*
Dinamarca	8
Países Bajos	8
Finlandia	5
Alemania	5
Gran Bretaña	5
Bélgica	4
Suecia	4
Rep.Checa	3
Noruega	3
Austria	2
Grecia	2
España	3
Estonia	1
Francia	1
Italia	1
Letonia	1
Polonia	1
Eslovaquia	1

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 los usuarios 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 identificación proactiva de necesidades sociales y situaciones de vulnerabilidad**. Desde la protección infantil en Dinamarca hasta la predicción de pobreza energética en los Países Bajos, estas aplicaciones buscan anticipar 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, dificultando 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, obteniendo un resultado coherente en gran medida con estudios previos (Tangi et al., 2022, p. 35).

3.1.1 Eficiencia administrativa y automatización

El marco de conceptos vinculados de una u otra forma a la eficiencia y la automatización creciente de la administración es el marco más notable en estos momentos. La investigación muestra una creciente externalización de tareas administrativas a la ciudadanía mediante autoservicios digitales, entre las que destaca la Automatización Robótica de Procesos (RPA), las comunicaciones automatizadas y clasificación de emails o mensajes, o la búsqueda y recuperación documental. Veamos los casos identificados en la literatura.

Automatización de procesos	RPA (Robotic Process Automation) en seguridad social	Finlandia, Suecia
	Optimización documental	Alemania
	Inteligencia colectiva para el cálculo de recursos en servicios sociales	España
	Sistemas de apoyo en la toma de decisiones (DSS, ADM)	Estonia, España
Procesamiento de lenguaje natural como apoyo	Emails certificados (Posta Elettronica Certificata)	Italia
	Chatbots para asistencia ciudadana	Bélgica, Finlandia, Reino Unido, España
	LocalLLMs para análisis de datos de bienestar infantil	Perron et al., 2024

RPA o Automatización Robótica de Procesos/ Robot Process Automation (Suecia, Finlandia). A menudo, los estudios hablan de *Robotic Process Automation* o RPA (Germundsson et al., 2024; Hutter & Hutter, 2021). En Finlandia, se automatizaron aproximadamente 500.000 decisiones anuales de ayudas y subsidios utilizando RPA, aunque enfrentó cuestionamientos del Defensor del Pueblo sobre su base legal (Väänänen, 2021). En Suecia existe, por otro lado, una automatización municipal relativamente masiva y aceptada por la sociedad en términos generales (Germundsson et al., 2024; Ranerup & Svensson, 2024).

Búsqueda de documentos (Alemania). Desarrollo del "GerPS-FIM-Microverse" (Raupach et al., n.d.), una representación semántica estandarizada para servicios públicos que facilita la digitalización de más de 10.000 servicios administrativos, constituyendo un caso prometedor de estandarización a gran escala. Relacionado con estos usos, encontramos los más recientes usos de LLM[10] basados en RAG.

Sistema de identificación de Demandas, Problemas y Respuestas (DPR) automático en Cataluña (España). El Ayuntamiento de Barcelona, Área de Derechos Sociales, ha implantado un sistema de apoyo a la toma de decisiones basado en la inteligencia colectiva, que realiza diagnóstico y previsión de recursos en base al entrenamiento con un repositorio de entrevistas con profesionales. El DPR automático está integrado dentro del sistema de información de los centros de servicios sociales y 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 (m4social, 2024, p. 78).

Sistemas de apoyo a decisiones o DSS (Decision-Support Systems). Durante los últimos años, se ha dado un foco importante en la aplicación de datos (master data) y sistemas de apoyo a la toma de decisiones en las instituciones de seguridad social de todo el mundo (Germundsson et al., 2024; Marienfeldt, 2024; Ruggia-Frick, 2021). Hablaríamos de toma automática o automatizada de decisiones o ADM cuando el sistema tiene cierta autonomía para tomar una decisión: p.e., el sistema RISK danés fue un sistema predictivo diseñado para mejorar las evaluaciones de riesgo de los trabajadores sociales en protección infantil (Ratner & Thylstrup, 2024), altamente controvertido.

Elegibilidad para recibir prestaciones (Estonia, España). El Töötukassa estonio, el fondo de financiación para situaciones de desempleo, usó sistemas de ADM para verificar automáticamente, en diferentes bases de datos integradas, si la persona había rellenado correctamente la información y si era elegible; este sistema, basado en reglas, automatiza el 50% aproximadamente de todas las decisiones en el país (OCDE, 2024).

Un sistema simular está implementado en el Departamento de Derechos Sociales de la Generalitat de Cataluña, donde desde diciembre de 2018 está en uso un motor de reglas de prestaciones sociales, dirigido a verificar si un ciudadano o entidad cumple los criterios para recibir una prestación social. A fecha de enero de 2024, se habían identificado cinco algoritmos desarrollados (m4social, 2024, p. 90): necesidades básicas, gastos del hogar, PNC

[10] Modelos de Lenguaje Grandes/Enormes; también los hay pequeños (Small Language Models) centrados en temáticas específicas. El ejemplo más conocido hoy en día es GPT de OpenAI.

jubilación, complemento de la PNC jubilación y prestación por nacimiento, acogida y adopción (familias).

Interfaces de toma de decisiones humano-IA. A diferencia de la década anterior, más recientemente, se tiende a introducir al humano en el "bucle" del Machine Learning para que pueda trazar y contrastar los datos que llevan a un resultado y que dicho resultado sea, además, explicable desde un punto de vista humano. Estos sistemas integran criterios algorítmicos con supervisión humana, como el robot Tengai en Suecia para procesos de reclutamiento en empleo público; en principio, estos procesos y su resultado serían imparciales y mediados por un humano (European Commission Joint Research Centre, 2020, Table 11).

Aunque pudiera parecer que hay cierto solapamiento de este grupo de casos con respecto a los anteriores de RPA, cabe considerarlo *cualitativamente* aparte, por separado, para poder introducir los sistemas de apoyo a la toma de decisiones (de humanos, bien sean responsables de la administración pública, técnicos, trabajadores sociales & c.) considerados éstos con una interacción humana activa a lo largo del proceso algorítmico.

En España encontramos wSocial, "una herramienta impulsada por el Departamento de Derechos Sociales y basada en la Inteligencia Artificial que, mediante el uso de palabras clave, permite detectar situaciones de vulnerabilidad y proponer intervenciones" (Fundació iSocial. Innovació en l'acció social, 2024).

Comunicaciones automatizadas (Italia). El Instituto Nacional de la Seguridad Social italiano implementó un sistema basado en IA para la clasificación y ordenación de correo electrónico certificado (Posta Elettronica Certificata) que clasifica automáticamente los miles de correos electrónicos certificados enviados al instituto cada día, permitiendo el reenvío inmediato a la oficina más adecuada (European Commission. Joint Research Centre., 2024, p. 22).

¿Qué es RAG? (Generación Aumentada por Recuperación)
RAG designa una técnica avanzada de prompting que combina un gran modelo del lenguaje (LLM, como GPT) con bases de conocimiento específicas del dominio, pero preparadas por cada organización en forma de chunks (trozos de información, en texto plano o en otro tipo de estructuras).
El objetivo es proporcionar información contextualizada en tiempo real, llegando a mostrar incluso una precisión del 91,7% en las respuestas generadas (Perron et al., 2025), lo cual abre un camino interesante de formación continua, pero también de análisis de información general a lo largo del ciclo de vida de un proyecto.

Chatbots y asistentes virtuales (Bélgica, Finlandia, España). El chatbot Ori del servicio de empleo belga (Office national de l'emploi – ONEM) es un caso de asistente virtual que puede responder una variedad de preguntas relacionadas con el desempleo, las bajas y pausas profesionales, incluyendo ayudas y elegibilidad. No toma decisiones por sí mismo. También encontramos otro chatbot, Kela-Kelpo/FPA-Folke, en Finlandia, con un enfoque similar: ayuda a los ciudadanos a completar solicitudes de subvenciones y prestaciones sociales, entendiendo finés, sueco e inglés (OCDE, 2024, pp. 22–23; Väänänen, 2021). Actualmente, en España se prueba un *recomendador* de prestaciones, plenamente conversacional, y otros chatbots para identificar necesidades expresadas en entrevistas, asistir a la identificación y validación para realizar trámites, o apoyar a la gestión de documentos a la ciudadanía. conversacional (m4social, 2024, p. 91 yss.)

3.1.2 *Sistemas de perfilado y segmentación*

Los sistemas de perfilado predictivo emergen como aplicaciones de la IA en políticas sociolaborales, empleando algoritmos para clasificar a beneficiarios según su probabilidad de necesitar intervenciones específicas. El perfilado es, hasta cierto punto, automático; esta situación supone riesgos para la transparencia algorítmica que se intentan solucionar con un paradigma "human-in-the-loop" (no exento, por lo demás, de riesgos, tal como veremos posteriormente en la sección de evaluación).

Perfilado y segmentación	Oportunidades de empleo	Austria
	Predicción de discapacidad laboral	Finlandia

Comparativamente, los sistemas de perfilado exhiben variaciones significativas en su diseño ético y orientación: se trata de un asunto complejo a nivel ético (Rachovitsa & Johann, 2022; Schmager et al., 2024; Züger & Asghari, 2023). Sirvan en esta sección a modo de ejemplo la monitorización de beneficiarios – actuales y potenciales - de servicios públicos como en Austria y Finlandia.

AMAS (Austria). El Servicio Público de Empleo austríaco implementó el "Arbeitsmarkt-Chancen Assistenzsystem" para calcular una "puntuación de oportunidad de integración" basada en historiales laborales y características personales, clasificando a los solicitantes de empleo en tres categorías según su probabilidad de reintegración laboral (Achterhold et al., 2025) Concretamente, se calculaba una puntuación de Oportunidad de Integración (IC) para cada individuo basada en su historial del mercado laboral y características personales.

Sistemas Predictivos de Discapacidad Laboral (Finlandia). En 2018, el Centro de Pensiones de Finlandia probó un algoritmo de aprendizaje automático que utilizaba una técnica estadística de autoaprendizaje para predecir si un individuo se jubilaría con pensión por discapacidad en dos años (Väänänen, 2021). El modelo alcanzó una precisión predictiva del 78%; este modelo de ML usaba datos socioeconómicos, de ingresos y prestaciones de 500.000 individuos.

3.1.3 *Sistemas de detección y respuesta temprana*

Vinculado al propio perfilado y segmentación, existe un subconjunto significativo de aplicaciones que se orienta hacia la identificación proactiva de necesidades o riesgos sociales, como se puede apreciar en los casos ilustrados a continuación.

Respuesta temprana	Servicios y ayudas proactivas	Bélgica
	Sistemas de alerta temprana para bienestar infantil	Dinamarca
	Predicción de la duración del desempleo	Eslovaquia
	Predicción de riesgo de pobreza y pobreza energética	Internacional, España, Países Bajos

Prestación de servicios "proactiva" (Bélgica). La ciudad de Gante desarrolló un sistema que utiliza datos existentes para otorgar proactivamente, especialmente a residentes vulnerables, descuentos automáticos en educación y cuidado infantil y otro tipo de ayudas (Kempeneer et al., 2024).

P**redicción de la duración del desempleo (Eslovaquia).** La Oficina Central de Trabajo, Asuntos Sociales y Familia de la República Eslovaca (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, mostrando una capacidad predictiva eficaz a partir de historial laboral, educación (cualificaciones) y 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 desarrollamos con posterioridad en las *secciones Implicaciones para las organizaciones* y *Conclusiones.*

Protección a la infancia (Dinamarca). Diversos proyectos algorítmicos intentaron identificar niños en riesgo, incluyendo un controvertido sistema en el municipio de Gladsaxe que experimentó con un sistema algorítmico para identificar niños en riesgo de abuso basándose en datos integrados desde diferentes sistemas, sociales, sanitarios y administrativos (European Commission Joint Research Centre, 2020). Este tipo de sistemas enfrentan y enfrentaron críticas considerables y fue judicializado debido al cuestionable enfoque y tratamiento de la privacidad (Cearns & Knox, 2024; Ratner & Schrøder, 2024; Ratner & Thylstrup, 2024).

Predicción de la pobreza (Internacional). La predicción de la pobreza es un campo en el cual el Machine Learning (ML) o 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.

Predicción de pobreza energética (España, Países Bajos). La Administración Abierta de Cataluña (AOC), en España, 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 propensos a error e ineficientes. 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, p. 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[11] para identificar factores predictivos

[11] Concretamente, se usó XGBoost (Extreme Gradient Boosting), que hace que muchos árboles de decisiones sean tomados en cuenta al mismo tiempo para tomar una decisión final sobre, en este caso, el riesgo de pobreza energética; para ello se identifican esos factores predictivos descritos (alquiler/propiedad, antigüedad, convivientes, etc.) y, posteriormente, el modelo los analiza en conjunto

como el valor de la vivienda y su estatus de propiedad, la antigüedad, número de personas por hogar y densidad de población (Dalla Longa et al., 2021).

3.1.4 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 beneficiarios finales, creando espacios de interacción antes inexistentes, facilitando 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; en una reciente revisión de la literatura (Babšek et al., 2025) documentan casos como CitizenLab en Bélgica o Civocrazy en Alemania. CitizenLab utiliza machine learning e IA generativa para procesar ideas del público y transformarlas en recomendaciones accionables para 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 impactos 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).

3.2 Aplicaciones de IA en la evaluación de políticas

Como se ha observado en el caso de la implementación de las políticas, en el ámbito evaluativo la aplicación de la IA cubre esencialmente la **automatización de determinadas tareas, de forma, por lo general, instrumental** y en el marco de una iteración "human-in-the-loop" – el *modelo más frecuente* o *predominante* - asemejándose a otros flujos de trabajo más generales.

Las tecnologías emergentes, como el aprendizaje automático, el procesamiento del lenguaje natural y el análisis de datos masivos o *Big Data*, tienen un importante potencial transformador, si bien no hay evidencia clara de su grado de adopción y uso en la profesión evaluadora.

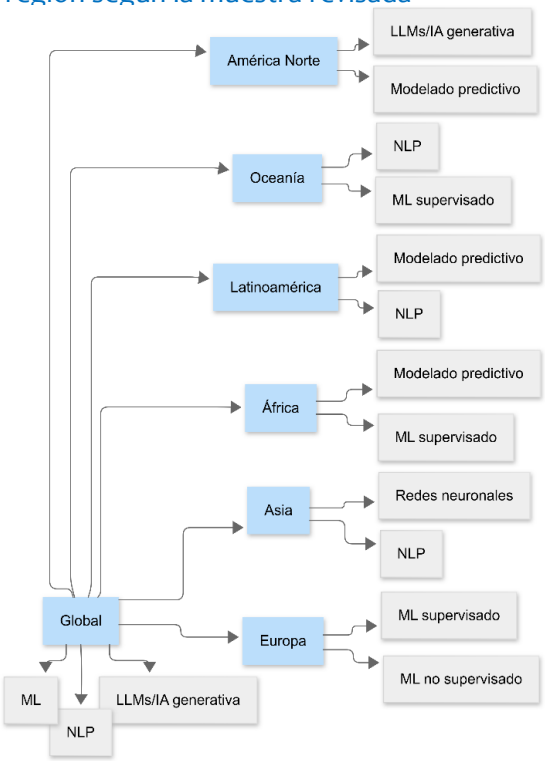
El análisis detallado de las técnicas empleadas en el marco evaluativo muestra un ecosistema tecnológico diversificado dentro del universo del aprendizaje-máquina (ML), incluyendo además el procesamiento de lenguaje natural (NLP), la IA generativa y el análisis visual.

Asimismo, se adoptan tecnologías de ML tanto supervisada como no supervisada, definidas respectivamente como:

- **Machine learning supervisado**, o SML: enseña a la máquina a codificar contenido de manera similar a como lo haría un humano.

- **Machine learning no supervisado**, o UML: Utiliza modelado de temas para agrupar conceptos relacionados, permitiendo procesar grandes volúmenes de texto de forma muy ágil.

Ilustración 6. Tecnologías IA adoptadas por región según la muestra revisada



Los **contextos geográficos** de la literatura revisada[12] evidencian una serie de patrones de adopción o de acceso a las tecnologías. Podríamos decir que, en esta revisión, aparece cierta especialización regional (sobre todo en Europa y EEUU) y convergencia metodológica global hacia lo híbrido, con múltiples tecnologías, que se adaptan al contexto, recursos y cultura.

A través de un análisis muy preliminar y descriptivo, con las cautelas requeridas por las limitaciones del método, y sin tener en cuenta variaciones intra-regionales, de una forma muy general se puede observar cierta predominancia de la IA generativa en América del Norte, de Machine Learning (tanto supervisado como no supervisado) en Europa, menciones a redes neurales con término de gran amplitud en Asia, y un foco en el análisis y procesamiento del lenguaje, y modelado predictivo en el Sur Global.

Para un análisis de los contextos territoriales, deberían analizarse en profundidad las variables específicas de cada país y los posibles factores influyentes. De forma general, cabe destacar que una parte importante de la literatura identificada en relación con países del Sur Global proviene del ámbito de la cooperación internacional al desarrollo y de organismos multilaterales de financiación, como el Banco Mundial o el sistema de Naciones Unidas. Estas instituciones han desempeñado un papel clave en la consolidación de marcos metodológicos y en la generación de evidencia empírica mediante evaluaciones sistemáticas, al contar con capacidades técnicas especializadas y recursos sostenidos.

Los **usos** encontrados se han agrupado en **cuatro** categorías principales:

1. Uso de datos masivos o Big Data.
2. Uso de simulaciones.
3. Uso de datos en tiempo real.
4. Uso de grandes modelos de lenguaje.

Los casos recogidos no permiten identificar la aplicación de la IA unívocamente en una u otra fase específica de la evaluación, sino acompañan **trasversalmente** todas las etapas del

[12] Las limitaciones subyacentes a la propia revisión se discuten en la sección 5.2Limitaciones del estudio, siendo conscientes de que pudieran existir resultados en otros idiomas y otros soportes que no habrían sido tenidos en cuenta y podrían sesgar esta interpretación de la adopción regional.

proceso evaluativo, centrándose especialmente en el **uso de los datos**, tanto para la fase de definición como de operacionalización y síntesis. Los casos ilustrados, en este sentido, ponen de manifiesto las oportunidades abiertas por la IA a lo largo de todo el ciclo de vida de los datos, desde la extracción y minería de datos masivos y simulaciones, a su procesado y análisis.

Del mismo modo, los grandes modelos de lenguaje (LLM) sirven transversalmente los flujos de trabajo **a lo largo del proceso evaluativo**, desde el diseño de los términos de referencia (*encargo*) a la redacción de informes y visualizaciones (*informe y uso*).

3.2.1 Uso de datos masivos o Big Data

El Machine Learning puede transformar la evaluación – y más allá, las capacidades – suponiendo una importante oportunidad de transformación (Hasan Chy & Nana Buadi, 2024).

La posibilidad de alcanzar grandes volúmenes de datos procedentes de una amplia pluralidad de fuentes heterogéneas y la capacidad de automatizar su procesado está transformando los procesos analíticos tradicionales, y con ellos la manera de abordar el manejo de los datos a lo largo del proceso evaluativo.

En la planificación y diseño (fases 1-3 del proceso evaluativo), la IA y los datos masivos pueden funcionar como herramientas de investigación preliminar y estructuración. Nos permiten investigar antecedentes, teorías del cambio, revisar y definir indicadores de desempeño, e incluso redactar términos de referencia (Jacob, 2025). Asimismo, podemos simular escenarios para estimar resultados o riesgos de diferentes acciones (Jacob, 2025), y realizar síntesis más rápidas y completas de los datos (Franzen, S. et al., 2022). También facilita la creación de perspectivas diacrónicas (longitudinales), valiosas para evaluar la sostenibilidad (Bohni Nielsen et al., 2024; York, 2024).

Por su parte, las nuevas herramientas computacionales permiten integrar y analizar información heterogénea de múltiples fuentes, transformando tanto los análisis cuantitativos como cualitativos. En el ámbito cuantitativo, facilitan evaluaciones contrafactuales y modelados causales que van más allá de la correlación para establecer relaciones de causalidad. Por otro lado, en el análisis cualitativo, las técnicas de Procesamiento de Lenguaje Natural (NLP) y los algoritmos de aprendizaje automático revolucionan el análisis de contenidos textuales, reduciendo los tiempos de procesamiento mientras mantienen niveles de precisión interpretativa.

3.2.1.1 Datos administrativos masivos (microdatos, Big Data)

El uso de datos administrativos masivos tiene una larga historia detrás (Crato & Paruolo, 2019a). La aplicación de tecnologías analíticas en la evaluación de políticas de empleo es particularmente relevante. El 68% de los estudios de evaluación de políticas activas de empleo en Europa utilizaban ya en 2019 datos administrativos procesados mediante tecnologías avanzadas (d'Hombres & Santangelo, 2019).

Actualmente, nos encontramos en un contexto marcado por una creciente disponibilidad de datos heterogéneos, de diversas fuentes y formatos. Las herramientas de Inteligencia

Artificial se pueden usar para limpiar, transformar y analizar datos, lo que reduce significativamente el tiempo necesario para estas tareas (Jacob, 2025). Además, la gran variedad de datos exige el desarrollo de metodologías analíticas sofisticadas y requiere formas innovadoras y prácticas para procesar y analizar esa información (Bamberger & York, 2020; Hasan Chy & Nana Buadi, 2024).

De este modo, la confluencia de la proliferación de fuentes de datos heterogéneas, el desarrollo de metodologías analísticas sofisticadas y la integración de estas capacidades, están transformando los procesos evaluativos en entornos complejos (Franzen, S. et al., 2022; Schimpf et al., 2021).

Datos masivos para evaluaciones contrafactuales (Portugal, Países Bajos). La combinación de métodos estadísticos avanzados, datos masivos e IA permite agilizar la realización de evaluaciones contrafactuales con datos administrativos, que a través del uso de grandes volúmenes de datos (datos masivos o Big Data) permiten determinar con mayor precisión si las políticas realmente funcionan y para quién quién (Bilbao-Goyoaga, 2023; Bohni Nielsen et al., 2024; Crato & Paruolo, 2019b). En este sentido, la integración de tecnologías con datos administrativos permite desarrollar sistemas muy sofisticados, integrando diversos conjuntos independientes de datos. A diferencia de los datos agregados, los microdatos – i.e., la información individual sobre personas o empresas – nos permiten determinar el impacto de una política (Crato & Paruolo, 2019a). Los dos casos reportados apuntan a la aplicación de datos masivos en el marco de evaluaciones contrafactuales. En el primero, reportado en Portugal, se han aplicado técnicas de regresión discontinua con datos administrativos longitudinales para evaluar programas de activación laboral en Portugal (Martins, 2019). En los Países Bajos, por otro lado, se han podido aplicar modelos de panel con efectos fijos para examinar la relación entre el patrimonio inmobiliario negativo y la movilidad laboral a través de la integración de 17 conjuntos independientes de datos administrativos (Morescalchi et al., 2019).

Análisis cuasi-experimentales y segmentados. Las capacidades computacionales están permitiendo el acceso a metodologías altamente complejas, tales como la Precision Analytics que combina Big Data y Machine Learning (ML) para realizar análisis cuasi-experimentales y segmentados (Bohni Nielsen et al., 2024, p. 46; York & Bamberger, 2020). Otro ejemplo es el del Modelado Causal Estructural o SCM (Bohni Nielsen et al., 2024, pp. 219–241), que busca explícitamente distinguir correlación de causalidad y entender qué funciona para quién. Mientras que muchos análisis estadísticos solo pueden determinar que dos variables están relacionadas, el modelado causal busca establecer si una variable realmente causa cambios en otra y en qué medida. Estos métodos usan datos de programas ya implementados, pero integran el conocimiento del experto humano "en el bucle", que aplica métodos de ML para construir modelos predictivos e inferenciales. El Causal Mapping es otra herramienta metodológica que se beneficia de estas capacidades (Powell et al., 2024).

Lo que distingue a estos enfoques es su capacidad para identificar relaciones causales y específicas para diferentes segmentos de población y determinar, por ejemplo, para qué grupos funciona mejor una intervención, qué elementos causan los resultados que vemos, o cómo ajustar las intervenciones.

3.2.1.2 Análisis cualitativo

Además de patrones y simulaciones matemáticas, se han identificado aplicaciones de IA también para el análisis cualitativo, a través del Procesamiento del Lenguaje Natural (NLP) y algoritmos de aprendizaje automático.

NLP y algoritmos de aprendizaje automático. El Procesamiento de Lenguaje Natural (NLP) hace posible y viable la tarea de analizar automáticamente grandes volúmenes de texto, extrayendo información valiosa de documentos, informes, comentarios y otras fuentes textuales que, en el pasado, requerirían meses de trabajo (Head et al., 2023; Tilton et al., 2023). Estas técnicas también se aplican para evaluar la calidad de programas o extraer información clave de documentos de proyecto, como en la práctica del IEG del Banco Mundial (Ziulu et al., 2024)y formulando sugerencias para acciones futuras (Jacob, 2025; MacArthur et al., 2025).

Análisis de contenido mediante algoritmos de aprendizaje automático. La IA puede automatizar el análisis de contenido mediante el uso de algoritmos de aprendizaje automático para procesar y clasificar grandes conjuntos de datos (Bohni Nielsen et al., 2024; Franzen, S. et al., 2022; Jacob, 2025). El caso de CoLoop en 2023 ejemplifica este potencial, logrando identificar en apenas 10-12 segundos los mismos temas generales en transcripciones que investigadores humanos habían extraído tras docenas de horas de codificación manual, manteniendo niveles comparables de precisión interpretativa en la evaluación de datos cualitativos (Jacob, 2025).

Los modelos de aprendizaje automático supervisados se pueden entrenar para que reconozcan temas o patrones específicos (Franzen, S. et al., 2022), con gran precisión. Ese fue el caso de Doing Business (Ziulu et al., 2024), en el que se han extraído datos textuales para la evaluación, aplicando análisis de texto y aprendizaje automático supervisado para identificación y evaluación de proyectos del Banco Mundial. Este modelo, además de poner de relieve la importancia de la interdisciplinariedad en la evaluación, nos señala un aspecto muy importante: la sostenibilidad de los modelos – que se demuestra con su transferibilidad - pues el modelo, bien diseñado, pudo ser re-usado con pequeños ajustes, lo que compensa los recursos inicialmente invertidos en su desarrollo.

3.2.2 Uso de simulaciones

Además del uso de datos masivos, otro campo emergente en la intersección entre Inteligencia Artificial y evaluación de políticas públicas es el de la simulación computacional. Estas técnicas permiten explorar escenarios alternativos, analizar dinámicas complejas y estimar impactos en contextos donde los enfoques tradicionales enfrentan limitaciones.

Existen antecedentes del potencial de las simulaciones computarizadas, tales como redes bayesianas, simulaciones de Monte Carlo, dinámica de sistemas, modelación basada en agentes en evaluación del diseño de políticas públicas, y experimentos bayesianos, y evaluación de generalizabilidad en impacto (Rodríguez Cano, 2018).

Más recientemente, se ha documentado el desarrollo de una herramienta de evaluación basada en la simulación de escenarios (Case-Based Scenario Simulation, o CBSS) en sistemas

complejos, que utiliza simulaciones de Monte Carlo y aprendizaje automático no supervisado (UML) para identificar patrones, corroborar resultados y simular escenarios alternativos para explorar contrafactuales, y plantear proyecciones futuras (Schimpf et al., 2021).

En el caso de España - y de especial interés, por el ámbito temático de este estudio, pues está aplicado al Ingreso Mínimo Vital - se ha evaluado el impacto del IMV con metodologías de control sintético basadas en aprendizaje automático supervisado (SML), simulando una España sintética (sin IMV), a modo de escenario contrafactual para aislar el impacto causal del programa comparando la situación real con la simulada (Bilbao-Goyoaga, 2023)

3.2.3 Uso de datos en tiempo real

La transición hacia sistemas de evaluación en tiempo real representa un cambio fundamental respecto a los modelos tradicionales. Esta evolución puede entenderse como un proceso gradual a través de tres "olas" tecnológicas (Bohni Nielsen et al., 2024, p. 273).

- La primera ola se caracterizó por la ampliación de fuentes de datos, incorporando información geoespacial, grandes registros administrativos y datos recopilados mediante dispositivos móviles. En este primer momento, las metodologías subyacentes no fueron sustancialmente modificadas.
- Posteriormente, la segunda ola, introdujo nuevas tecnologías para la captura y análisis de datos, como técnicas basadas en IA y aprendizaje automático, Internet de las cosas, satélites y drones; no se quedaron tan sólo en una dimensión instrumental, pues abrió posibilidades y modificó de forma efectiva la metodología del análisis.
- La tercera ola, parcialmente superpuesta con la anterior, integró tecnologías avanzadas para la captura, almacenamiento y procesamiento de datos a gran escala, y en tiempo real: esto sí supone un cambio mayor en la concepción de la evaluación misma.

Los sistemas modernos de monitoreo proporcionan un flujo de datos recurrente, que puede posibilitar el seguimiento continuo durante la implementación de programas y políticas (Bohni Nielsen et al., 2024, pp. 275–277) y su ajuste en tiempo real – e.g., mediante el análisis de los datos y la generación de puntuaciones de riesgo (Bamberger & York, 2020; Buttow, 2024, p. 15; Jacob, 2025).

Esta adaptabilidad en tiempo real hace posible la incorporación de la evaluación en cada fase del ciclo político, sustituyendo la noción de evaluación como etapa final: podemos realizar evaluaciones que se entregan en tiempo real desde el inicio mismo de la implementación, dando lugar a una evaluación continua en políticas públicas (Valle-Cruz et al., 2020).

De este modo, la aplicación de la IA da un vuelco a la *ubicación* de la evaluación en el ciclo de las políticas, al absorberla y distribuirla en las distintas fases de un proceso continuo (Carlizzi & Quattrone, 2023). La evaluación se "desborda" en todo el ciclo de la política pública: los datos masivos son recogidos y categorizados, a la vez que analizados, durante la configuración de las prioridades en la agenda pública; el modelado predictivo y los sistemas de apoyo a la toma de decisiones (DSS) se integran en el proceso de elección de alternativas; durante la fase de implementación se puede customizar la "Micro-política" con herramientas de eDemocracy, chatbots, análisis de sentimiento, etc. (Carlizzi & Quattrone, 2023).

No obstante, como señala Valle-Cruz (Valle-Cruz et al., 2020), este enfoque puede conducir a una falta de teoría sustantiva para comprender las implicaciones de los datos, como se verá en el apartado dedicado a las limitaciones y riesgos en la aplicación de la IA en la evaluación.

3.2.4 Uso de grandes modelos de lenguaje (LLM)

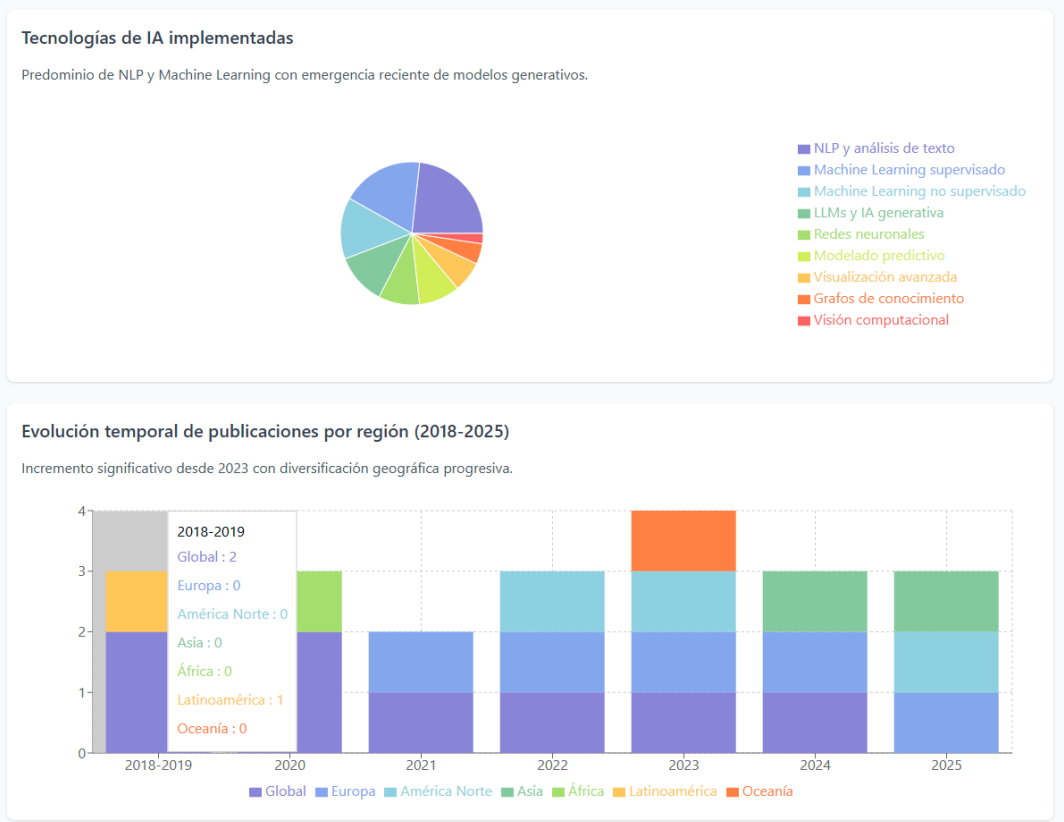
Los grandes modelos de lenguaje (LLM) ofrecen apoyo en diversas tareas del proceso evaluativo.

Planificación y diseño. Puede asistir en la planificación (investigar antecedentes, definir indicadores, redactar ToRs), formular preguntas para entrevistas, generar visualizaciones de resultados (Jacob, 2025). Aunque no exenta de problemas (Ferretti, 2023; Head et al., 2023; Kates & Wilson, 2023), la IA Generativa mejora la capacidad de comprensión y análisis del contexto, reforzando el proceso evaluativo global.

Redacción. El potencial e impacto de la IA generativa en tareas de redacción puede ser considerable, siendo capaz de automatizar y acelerar cualesquiera generación de contenido, especialmente textual (pero no únicamente), lo que apuntaría a la automatización de partes significativas de la escritura de informes (Head et al., 2023). Asimismo, puede emplearse para la generación de documentos altamente protocolizados y rígidos, que forman parte de las áreas de competencia de fácil sustitución (Jacob, 2025; Mason, 2023), y se emplea ya hoy en día para la generación de propuestas y memorias técnicas (Kates & Wilson, 2023) u otra documentación técnica formalizada, términos de referencias, marcos lógicos si sólo actúan a modo de precursores de informes, o traducciones multilingües o para múltiples audiencias (Ferretti, 2023).

Visualización y presentación. La incorporación de la IA no solo transforma la recolección y análisis de datos, sino también su presentación a las partes interesadas. La IA puede presentar los resultados de las evaluaciones en varios formatos, como informes, vídeos y podcasts, adaptados a diferentes públicos (Jacob, 2025). También puede traducir a varios idiomas (Ferretti, 2023) y generar visualizaciones interactivas, como paneles y gráficos (Jacob, 2025). Estas innovaciones no son menores: contribuyen, en diferentes grados según su contexto, a hacer que todo el proceso, desde la recolección de datos hasta su comunicación, sea más accesible e inclusivo.

Ilustración 7. Ejemplo de IA generativa para la visualización de datos



(Claude 3.7 Sonnet Extended) Prompt: "¿Puedes hacer un panel de visualización atractivo para enseñarlo a las partes interesadas?".

3.3 Implicaciones para las organizaciones evaluadoras

La capacidad de una organización para adquirir, interpretar y aplicar conocimiento para adaptarse y mejorar – i.e. el aprendizaje organizacional – es un factor fundamental en toda institución evaluativa y/o investigadora. La irrupción de la IA está catalizando una profunda transformación en este ámbito, en lo que refiere a la conciencia situacional como a la memoria institucional y redefiniendo la colaboración y entre qué-quién se da.

3.3.1 Conciencia situacional y memoria organizacional

Por un lado, la IA permite a las organizaciones desarrollar una conciencia situacional más aguda, tanto a nivel operativo como estratégico. Al procesar grandes volúmenes de información del entorno, de las propias operaciones y de las interacciones con la comunidad, podemos detectar tendencias emergentes, anticipar cambios y facilitar una adaptación más ágil, más rápida y eficaz (Dwivedi et al., 2021).

Por otro lado, la IA proporciona en sí, entre otras cosas, un repositorio de "memoria organizacional" que puede almacenar lecciones aprendidas o buenas prácticas, fomentar el aprendizaje y dar apoyo en la mejora continua de las organizaciones (Elevati, 2025). Un ejemplo destacado es el uso de análisis automatizado de texto y visualización para detectar

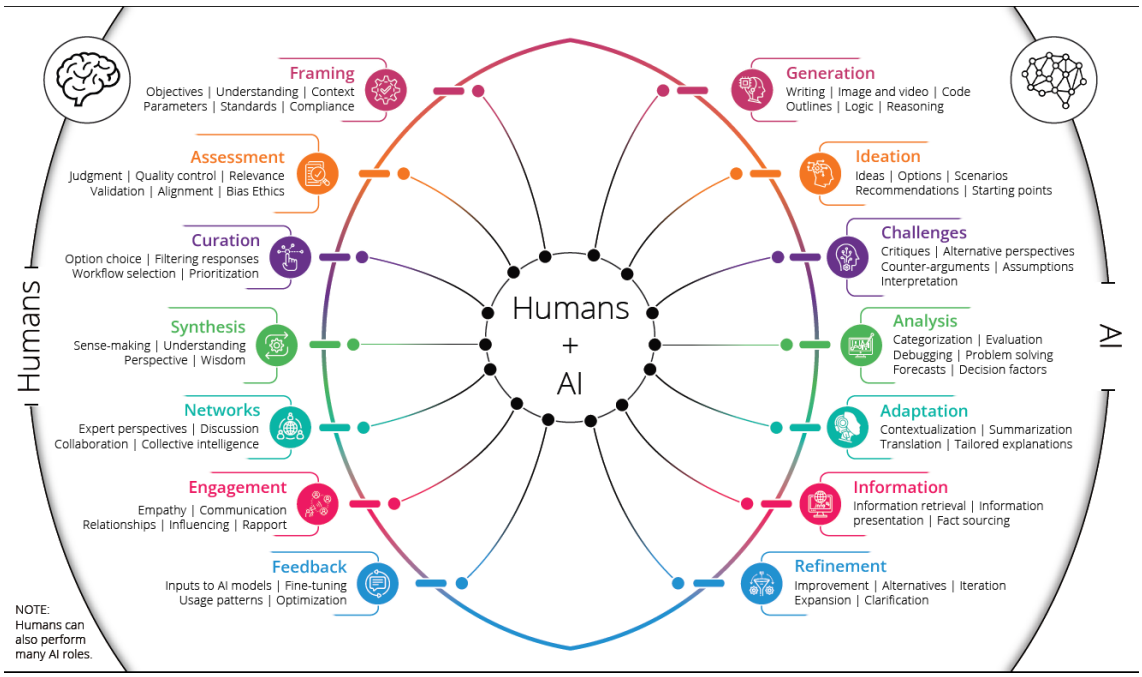
tendencias y cambios temáticos en 30 años de informes de desempeño en Israel, creando una memoria institucional dinámica y accesible (Raveh et al., 2020).

Además, esta capacidad de reutilizar y construir sobre conocimiento pasado puede hacer más sostenibles las inversiones en modelos analíticos complejos (Ziulu et al., 2024) sobre todo considerando su transferibilidad cuando la programación del modelo sigue las buenas prácticas.

3.3.2 *Interacción humano-máquina*

Las formas de colaboración e interacción también se están transformando con el avance de las aplicaciones de IA en las organizaciones. La integración efectiva de la IA en la evaluación rara vez implica una automatización total. En su lugar, emerge un modelo de colaboración sinérgica entre la inteligencia humana y la artificial (Elevati, 2025), a menudo descrito con el concepto "human-in-the-loop". Este enfoque implica que los evaluadores trabajen junto a los algoritmos, ya sea supervisando el proceso general ("human-on-the-loop") o interviniendo en decisiones específicas ("human-in-the-loop"), como se muestra en la ilustración, asegurando que el juicio crítico, la comprensión contextual y la responsabilidad ética permanezcan en el centro del proceso evaluativo (Bohni Nielsen et al., 2024, p. 248).

Ilustración 8. Flujos de trabajo Humano-IA



Fuente: (Harris & Dawson, 2025).

Mantener la fluidez tecnológica se vuelve indispensable a medida que los formatos legibles por máquina y las APIs se integran en el ecosistema de datos (Bohni Nielsen et al., 2024). Este enfoque permitiría, en un escenario ideal, que la IA se encargue de tareas rutinarias o computacionalmente intensivas, liberando a los evaluadores para centrarse en la interpretación, el razonamiento complejo y la interacción humana (Elevati, 2025) y en co-

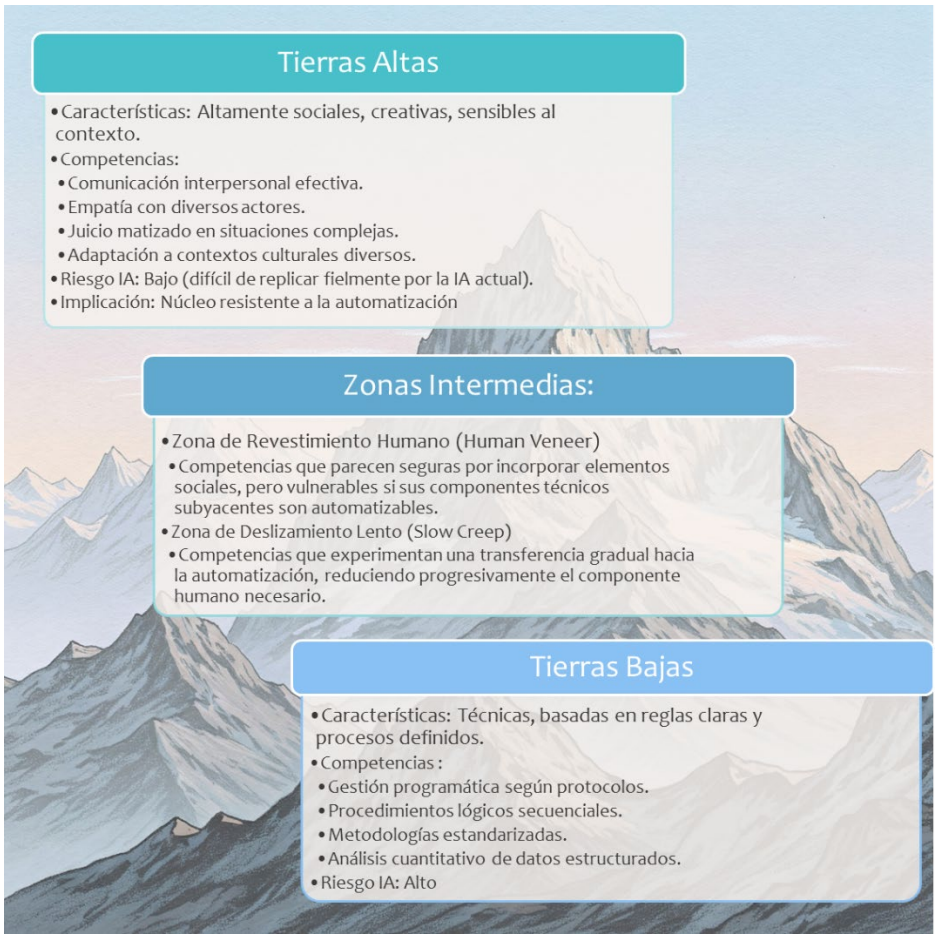
diseñar áreas donde se pueden lograr mejoras en eficiencia, eficacia y valor público (Babšek et al., 2025).

3.3.4 Competencias y riesgos de sustitución

Este nuevo panorama impacta directamente en las competencias requeridas para los evaluadores y, por extensión, en la capacidad de aprendizaje de las organizaciones (Mason, 2023). La tendencia apunta a la necesidad de fortalecer las competencias interpersonales, el pensamiento crítico y la adaptabilidad contextual, al tiempo que se desarrolle una sólida alfabetización en IA y ciencia de datos (Bouyousfi & Ouedraogo, 2024; Picciotto, 2020).

En el modelo desarrollado por Mason para este análisis, vemos como frente a competencias que presentan mayor riesgo de automatización y sustitución, como tareas técnicas, rutinarias, de gestión (las "tierras bajas", o "zonas de peligro"), los aspectos del trabajo evaluativo que implican interacción social, respuesta a contextos específicos, imprevisibilidad y creatividad (las "tierras altas") constituirán el núcleo sostenible de la profesión frente al avance de la IA (Mason, 2023). Mason recomienda enfatizar estas competencias tanto en la práctica cotidiana como en la formación de nuevos evaluadores, parafraseando el consejo de Max Tegmark: "desarrollar habilidades evaluativas en las que las máquinas son deficientes: aquellas que involucran personas, imprevisibilidad y creatividad".

Ilustración 9. Marco de competencias y riesgo de automatización



Fuente: Elaboración propia, a partir de (Mason, 2023) Ilustración de fondo generada con Gemini 2.0 Flash.

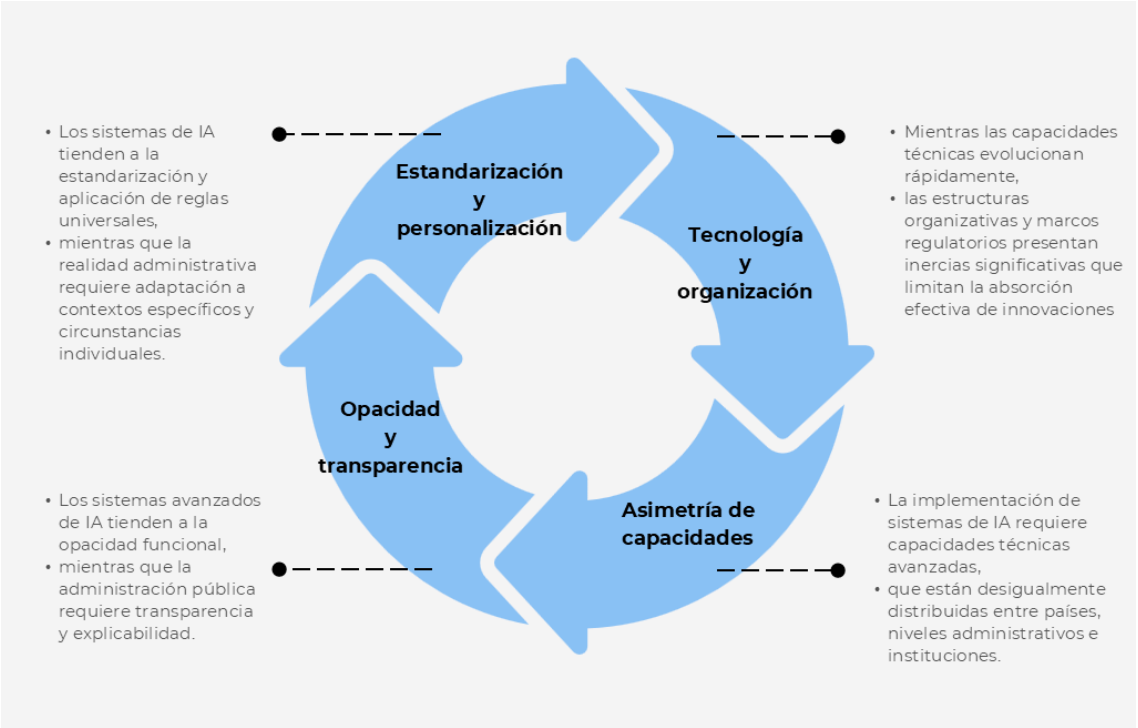
En la misma línea Ferretti, quien identifica en la automatización de determinados procesos la ventaja de liberar tiempo para los aspectos más creativos y contextuales del trabajo evaluativo, precisamente aquellos que Mason identifica como menos susceptibles a la automatización (Ferretti, 2023).

Las organizaciones evaluadoras necesitan, consecuentemente, fomentar activamente esta adaptación, integrando conocimientos sobre IA en la formación continua y los planes de estudio (Tilton et al., 2023), y promoviendo una cultura de la evaluación crítica sobre los propios efectos de la IA en la práctica evaluativa (Ferretti, 2023; Kates & Wilson, 2023).

4. Tendencias y retos

Del análisis de los artículos seleccionados, emergen tensiones y complementariedades significativas en esa complejidad que implica la transformación digital en general, y más aún, en el ámbito público; 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.

Ilustración 10. Tensiones en la aplicación de la IA



Fuente: elaboración propia a partir de distintas fuentes (Babšek et al., 2025; Bajgar & Criscuolo, 2019; Bohni Nielsen et al., 2024; Bouyousfi & Ouedraogo, 2024; Jacob, 2025; Kates & Wilson, 2023; Mason, 2023; Saura & Debasa, 2022)

4.1 Principales ventajas

La literatura revisada muestra, de hecho, cómo la aplicación de IA ha generado un incremento de eficiencia y reducción de costes y tiempo (m4social, 2024; OCDE, 2024; Perron et al., 2024; Ranerup & Svensson, 2024; Väänänen, 2021), procesos y asignación de recursos (Marienfeldt, 2024; OCDE, 2024; Schmager et al., 2024; Väänänen, 2021) e incluso en calidad, accesibilidad y personalización de servicios (Lee-Archer & ILO Regional Office for Asia and the Pacific, 2023; m4social, 2024).

Además, la IA nos permite - en cierta medida - la prestación proactiva de servicios e intervención temprana (Kempeneer et al., 2024; Lehtiniemi, 2024; m4social, 2024; OCDE,

2024; Ratner & Schrøder, 2024) y, con las salvaguardas adecuadas, podría fortalecer el apoyo específico a colectivos 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 seguridad social para mejorar la prestación de servicios a ciudadanos y empleadores en Bélgica (Lee-Archer & ILO Regional Office for Asia and the Pacific, 2023) o de arquitectura digital interconectada en las aplicaciones de Inteligencia Artificial y toma de decisiones automatizada en la Institución de Seguridad Social de Finlandia (Väänänen, 2021) - muy dependiente, a su vez, de la confianza de la ciudadanía en el gobierno. Estos casos, junto a los estudiados en Cataluña, aparecen como posibilidades en materia de inter-operabilidad de datos, pero a la vez resaltan las dificultades que existen para poder analizar de forma coordinada y compartir, no ya datos, sino tecnologías.

Ilustración 11 Principales tendencias

Fuente: elaboración propia a partir de diversas fuentes (Bouyousfi & Ouedraogo, 2024; Elevati, 2025; Hasan Chy & Nana Buadi, 2024; Marienfeldt, 2024; Martins, 2019; Ratner & Schrøder, 2024; Schmager et al., 2024).

4.2 Retos y desafíos

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 (Minguijon & Serrano-Martinez, 2022). En lo referente a la privacidad y no-discriminación, se encuentra 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 & Johann, 2022; Schmager et al., 2024; Züger & Asghari, 2023) y se plantean una serie de problemáticas, tales como como la atribución de responsabilidad (riesgo moral), fragmentación y silos**,** cajas negras (falta de transparencia) y sesgo algorítmico (riesgo de

amplificar sesgos históricos, mencionando el preocupante sesgo étnico en detección de fraude).

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, señalando incorrectamente a muchas familias como potenciales defraudadoras, especialmente afectando a personas de bajos ingresos y 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 – ie., el efecto de riesgo moral (Lee-Archer & ILO Regional Office for Asia and the Pacific, 2023).

Sesgo algorítmico. Existe un riesgo, cuando los sistemas de IA utilizan datos históricos, de importar y amplificar sesgos de los cuales los usuarios podrían no ser conscientes (European Commission Joint Research Centre, 2020). Esto es, entre otras cosas, el sesgo algorítimico. De encontrarse, por ejemplo, un sesgo étnico en el modelo, deberían poderse identificar las relaciones que producían ese sesgo y tratar de reducirlas (European Commission. Joint Research Centre., 2024; Rachovitsa & Johann, 2022; Ratner & 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, dificultando entender cómo se toman las decisiones automatizadas. Esta opacidad puede erosionar la confianza y limitar la rendición de cuentas (European Commission. Joint Research Centre., 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, dificultando el escrutinio externo. Solo tras presiones se publicó una representación basada en regresión logística, revelando que atributos como género y nacionalidad impactaban negativamente en la puntuación de reintegración laboral, actuando como un sesgo algorítmico. Esta opacidad generó preocupaciones sobre la reproducción de discriminaciones históricas, haciendo de AMAS un ejemplo paradigmático de "caja negra algorítmica". Frente a ello, indicadores de equidad[13] y técnicas de mitigación de sesgos[14] pueden permiten 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, neutralidad e imparcialidad, propios del mandato del servicio público, y transparencia, rendición de cuentas y posibilidad de objeción,

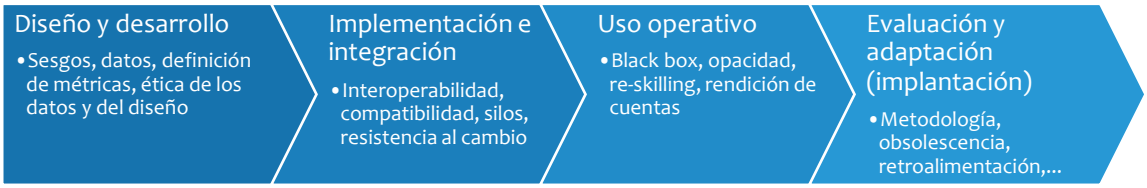
[13] Medidas estadísticas que evalúan la equidad de un modelo algorítmico observando cómo se distribuyen los resultados y errores entre distintos grupos definidos por atributos sensibles, como el género. Estas métricas permiten detectar disparidades sistemáticas en las decisiones automatizadas.
[14] Técnicas de mitigación de sesgos como: reponderación de datos, representaciones justas, regresión logística con restricciones de equidad y post-procesamiento por paridad de oportunidades.

especialmente en decisiones que afectan directamente la vida de los ciudadanos (Schmager et al., 2024).

Fragmentación de datos y silos organizacionales. En múltiples casos analizados, se ha observado cómo diferentes 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. (European Commission. Joint Research Centre., 2024).

Incompatibilidad e interoperabilidad. A nivel más general, la imagen en conjunto de los servicios públicos electrónicos sigue siendo algo fragmentada debido al gran número de instituciones involucradas (Lee-Archer & ILO Regional Office for Asia and the Pacific, 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 ciudadanos (Cearns & Knox, 2024).

Ilustración 12. Proceso de integración de tecnologías y barreras



4.3 Reflexiones desde la práctica evaluativa

Como se ha visto, la tendencia dominante en la literatura aboga por ver la IA como un complemento, como "una herramienta", y no un sustituto, ni de los métodos tradicionales, ni del juicio experto del evaluador (Bouyousfi & Ouedraogo, 2024; Schimpf et al., 2021) . El enfoque "human-in-the-loop" se presenta como dominante y, hasta cierto punto, fundamental (Elevati, 2025), asumiendo que la IA – cualesquiera IAs - apoya tareas específicas y "mecánicas (e.g., análisis de grandes volúmenes de datos, identificación de patrones) y que la interpretación final, el juicio evaluativo, la consideración del contexto y la validación ética recaen en el evaluador humano. No obstante, esto no está exento de dificultades, ni es la solución finalista que aparenta ser: existe el riesgo de que se convierta en un "cascarón procedimental vacío" que legitime el uso de algoritmos problemáticos sin garantizar una protección real de derechos y, además, la capacidad humana para supervisar eficazmente decisiones algorítmicas complejas es limitada (Bohni Nielsen et al., 2024, p. 252). Se suma a esto el riesgo del "sesgo de automatización", donde los humanos tienden a confiar excesivamente en los resultados algorítmicos (Elevati, 2025).

En ese sentido, las inquietudes desde la práctica evaluativa se alinean a los principales retos destacados anteriormente, especialmente en lo relacionado con la calidad y representatividad de los datos de entrenamiento, la explicabilidad e interpretación de las cajas negras, los problemas de veracidad, privacidad y protección de los datos.

Calidad, sesgo y representatividad. Los estudios detectan una potencial perpetuación de sesgos históricos y exclusiones epistémicas. El sesgo es un aspecto crítico que ya hemos tratado a lo largo de este informe (Bouyousfi & Ouedraogo, 2024; Ferretti, 2023; Head et al., 2023; Jacob, 2025; Stern, 2020; Tilton et al., 2023), pero, ¿hasta qué punto el propio sesgo podría ser perpetuado por el uso de tecnologías emergentes en la evaluación?

La validez de los modelos de IA depende intrínsecamente de la calidad y representatividad de los datos de entrenamiento. Existe un riesgo significativo de perpetuar o amplificar sesgos históricos y desigualdades si los datos reflejan discriminaciones pasadas o sub-representan a ciertos grupos (Bouyousfi & Ouedraogo, 2024; Elevati, 2025; Hasan Chy & Nana Buadi, 2024). La falta de representatividad puede derivar también de factores contextuales, como el fenómeno de "non take-up" en programas sociales, que introduce sesgos sistemáticos no atribuibles a la calidad intrínseca del dato (Bilbao-Goyoaga, 2023; Bouyousfi & Ouedraogo, 2024).

Explicabilidad e interpretación en las cajas negras. Los modelos altamente complejos como las redes neuronales profundas o los grandes modelos de lenguaje (LLM) a menudo funcionan como "cajas negras", dificultando la comprensión de cómo llegan a sus conclusiones (Hasan Chy & Nana Buadi, 2024; Jacob, 2025; Kates & Wilson, 2023). Esta opacidad choca con los principios de transparencia y rendición de cuentas cruciales en la evaluación de políticas públicas, y nos fuerza a elegir - y es un compromiso reconocido- entre la precisión predictiva (a menudo mayor en modelos opacos) y la interpretabilidad (Bohni Nielsen et al., 2024, p. 96). La emergencia de enfoques de IA Explicable (XAI) busca mitigar este problema, pero sigue siendo un desafío central.

Precisión factual y alucinaciones en la IA Generativa. Los modelos grandes de lenguaje tienen un riesgo importante de generar información fabricada, pero plausible, por su propio diseño (Ferretti, 2023; Head et al., 2023). Este problema de veracidad (Kates & Wilson, 2023) subraya la necesidad imperativa de validación humana

Privacidad y protección de los datos. El uso de grandes volúmenes de datos, a menudo de carácter personal o administrativo, suscita graves preocupaciones éticas sobre el consentimiento informado, la privacidad, la propiedad de los datos y el potencial uso indebido de información sensible; esto es especialmente delicado cuando hablamos de la prestación de servicios y la implementación e implantación de política social (Bamberger & York, 2020; Bouyousfi & Ouedraogo, 2024; Hasan Chy & Nana Buadi, 2024).

Por otro lado, la práctica evaluativa evoca desafíos propios de la profesión, desde la necesidad de adquirir competencias a nivel tanto organizacional como individual, hasta auténticos desafíos epistemológicos, que cuestionan la pluralidad metodológica, la transparencia, la equidad, la validación rigurosa.

Brecha de competencias. La brecha de competencias no es tan sólo un diagnóstico de carencias en el presente europeo: es una aseveración que plantea problemas entre organizaciones, capacidad financiera, y regiones. La implementación de IA conlleva costos significativos en infraestructura, software y personal especializado, en un entorno en el cual se suman la resistencia al cambio por parte de profesionales (Babšek et al., 2025; Bamberger & York, 2020), la escasez de evidencia sistemática sobre implementaciones exitosas y la falta de marcos regulatorios y conceptuales claros(Babšek et al., 2025; Hasan Chy & Nana Buadi, 2024).

Imperativo de cuantificación. De forma relacionada, parece surgir cierto imperativo de cuantificación que, por lo demás, no es nuevo. Si bien existen aplicaciones cualitativas (e.g., Procesamiento de Lenguaje Natural - NLP), el marco epistemológico subyacente a muchas aplicaciones de IA tiende a privilegiar lo cuantificable. Esto genera una tensión con el pluralismo metodológico valorado en evaluación y plantea interrogantes sobre la conmensurabilidad de fenómenos cualitativos y el lugar de fenómenos no cuantificables.

A pesar de las capacidades analíticas avanzadas, la IA no resuelve por sí misma los desafíos inherentes al establecimiento de inferencias causales robustas en la evaluación. No causalidad no es asunto de técnicas y tecnologías, sino de diseño experimental. Su uso instrumental no hace que deje de persistir la dificultad de aislar los efectos del programa de las características preexistentes de los participantes o de factores contextuales(Bajgar & Criscuolo, 2019; Bouyousfi & Ouedraogo, 2024; Schimpf et al., 2021; Stern, 2020).

Determinismo tecnológico, contexto, experticia y validación. Algunos autores destacan que la adquisición de una "tecnoestructura" avanzada no garantiza por sí sola mejores resultados si la autoridad se concentra en expertos técnicos sin contrapesos (Alexander, 2022). Se observa cierta tendencia persistente hacia el determinismo tecnológico en el marco de la relación entre la IA y la práctica de la evaluación. Si bien se reconocen los desafíos de implementación, los documentos presentan predominantemente una narrativa de progresión en la que la evaluación evoluciona inexorablemente desde los enfoques tradicionales hacia metodologías aumentadas o impulsadas por la IA. Esta contextualización podría obviar las contingencias contextuales de adaptación-apropiación de la tecnología y la trayectoria no-lineal en diferentes instituciones; tampoco da cuenta de efectos derivados de los errores de implementación. Presentar la adopción de IA como una progresión lineal e inevitable obviaría las contingencias contextuales, las trayectorias no lineales de adopción y los efectos imprevistos derivados de la implantación, pues el comprender el impacto de un sistema de IA requiere analizar no solo sus datos y modelo, sino también la infraestructura material, el contexto de despliegue y su interacción las personas, la sociedad, y las instituciones (Bohni Nielsen et al., 2024, p. 257).

La promesa de la IA solo podrá materializarse de forma responsable si se prioriza la transparencia, la equidad, la validación rigurosa. El juicio humano no sólo tiene que ser técnicamente informado, sino éticamente consciente en un marco de colaboración interdisciplinar robusta (Bamberger & York, 2020; Ferretti, 2023; Franzen, S. et al., 2022; Jacob, 2025; Mason, 2023; Schimpf et al., 2021).

5. Limitaciones

5.1. Limitaciones de la literatura analizada

La literatura científica analizada presenta varias limitaciones que deben considerarse.

En primer lugar, existen diferentes problemáticas en el diseño y evaluación de los sistemas. Muy a menudo, la publicación se limita a casos exitosos y la visión políticamente positiva de tales implementaciones; por otro lado, los datos utilizados para el análisis cruzado se limitan a registros y documentos publicados secundarios (Saura & Debasa, 2022). Un único ejemplo no es suficiente para realizar juicios de evaluación generales sobre el sistema y, por sesgo de publicación, hay sobrerrepresentación de experiencias exitosas. Experiencias además sujetas muy especialmente a cierta obsolescencia, pues el ritmo de los avances tecnológicos – de por sí rápido y enfocado en la eficiencia de las organizaciones desarrolladoras – se acelera más aun en tecnologías emergentes estratégicas, como es el caso de la IA y, desde 2022 – aunque, en realidad, 2017 – la GenAI. Esta dificultad plantea un problema para establecer comparativas longitudinales, por ejemplo, en los efectos de la implantación tecnológica (ie., de sistemas de IA) en política social, política pública o prestación de servicios.

En segundo lugar**, podría haber cierto tipo de problemas en captura de perspectivas múltiples**. Por ejemplo, las visiones y articulaciones de algo tan crucial como **qué son "los datos"** podría variar, especialmente, en qué deberían y qué pueden hacer (Cearns & Knox, 2024). En este sentido, es complejo integrar las perspectivas divergentes de tecnólogos, funcionarios públicos, ciudadanos y responsables políticos. La desagregación analítica, además, es insuficiente, debido al uso de datos administrativos pasados: la propia metodología puede hacer imposible realizar un análisis separado de las respuestas dadas por los grupos involucrados (cf. gerentes, políticos y trabajadores sociales) (Ranerup & Svensson, 2024).

En tercer lugar, pero muy vinculado a las anteriores cuestiones, la complejidad causal aparece como una ambición y un problema. Se pueden llegar a estudiar fenómenos altamente complejos, lo cual es una fortaleza pero también una problemática de implementación y de investigación (Dalla Longa et al., 2021). El **razonamiento predictivo** sobre cuestiones de alta complejidad es problemático por la misma razón, incluyendo la prevención de daños (Lehtiniemi, 2024).

Por otro lado, la mayoría de los estudios, que están sujetos a ese sesgo de publicación, **describen comportamientos atípicos, y no generalizables**; incluyendo cómo se aplican "soluciones manuales" en el manejo de las herramientas digitales para equilibrar las reglas universales, que a menudo son enfatizadas por las herramientas digitales de una manera rígida y restrictiva (Marienfeldt, 2024).

Los usos más avanzados y de alto impacto de la IA (desde el análisis predictivo hasta la toma de decisiones asistida por IA) han tenido **poca implementación** en el mundo real; menos aun, en el contexto de los servicios públicos (Schmager et al., 2024). Por ello, hay una distancia significativa entre las capacidades teóricas y las implementaciones prácticas: los enfoques existentes no siempre se pueden trasladar – o no directamente - al contexto del sector

público y deben adaptarse y analizarse individualmente (Schmager et al., 2024), contextualmente, socialmente y legalmente. También existe hoy un panorama ético-social cambiante: las mesetas éticas, la adaptación de los profesionales, y el rol de la confianza plantean un reenfoque del uso de IA, que se evidencia en cómo, con el tiempo, a los modelos algorítmicos se les concede menos agencia y poder, reflejando negociaciones continuas sobre la distribución humano-máquina de dicho poder, ya re-negociación del humano-profesional (Marienfeldt, 2024; Ratner & Schrøder, 2024) frente a sistemas automatizados.

5.2. Limitaciones del estudio

La metodología empleada en esta scoping review presenta un enfoque estructurado y teóricamente sólido para mapear la aplicación de la inteligencia artificial (IA) en políticas de empleo e inclusión social. No obstante, de manera similar a los diseños metodológicos protocolizados, tiene limitaciones y aspectos susceptibles de mejora o ampliación.

La **inclusión de motores de búsqueda generalistas** como Google, junto a bases de datos académicas es metodológicamente pertinente, aunque su efectividad depende críticamente de la especificidad de los algoritmos de búsqueda: los algoritmos de búsqueda deben captar, por otro lado, un **espectro amplio** de la literatura gris. Además, los motores generalistas **priorizan** resultados basándose en algoritmos complejos (SEO, relevancia percibida, historial del usuario, etc.) que no necesariamente coinciden con la **relevancia** académica o institucional.

Los **términos de búsqueda, aunque amplios, podrían no haber capturado la jerga específica**, y la indexación de contenido en los motores de búsqueda no es instantánea. Existe un lapso, variable, entre la publicación de un documento (especialmente informes institucionales o literatura gris en sitios web específicos) y su aparición efectiva en los resultados.

Además, la exclusión de artículos periodísticos y blogs - aunque justificada por la búsqueda de rigor empírico - podría omitir **señales** tempranas de innovación y debate público. Como todas las revisiones de literatura, esta scoping review se basa en evidencia documentada: i.e., iniciativas muy novedosas, experimentales o internas dentro de las administraciones que aún no han generado informes o publicaciones formales quedarán fuera del alcance.

Es importante resaltar que estas limitaciones metodológicas se ven acentuadas por las restricciones de tiempo y recursos disponibles para la realización del estudio, factores que inevitablemente condicionan el alcance y la profundidad del análisis realizado.

La literatura gris más reciente ofrece ejemplos significativos sobre la implementación de IA en servicios públicos de empleo (SPE). Estos estudios, de forma paralela a nuestra revisión, reflejan las tendencias observadas, como el enfoque inicial en la elaboración de perfiles y la orientación, el creciente interés en el emparejamiento (aunque con desafíos), y la exploración de tecnologías más avanzadas como los LLM y la GenAI.

El informe de la **Comisión Europea** (European Commission. Directorate General for Employment, Social Affairs and Inclusion & ICF., 2025) destaca el papel crucial del movimiento Big Data a mediados de la década de 2010 como impulsor de la atención hacia la IA, identificando su potencial para mejorar la prestación de servicios, aumentar la eficacia de

los procesos mediante emparejamiento y elaboración de perfiles más precisos, optimizar la eficiencia mediante herramientas de apoyo a la toma de decisiones, y ampliar el uso de datos y análisis. Los datos masivos no son propiamente IA, pero sí son su insumo, tal como refleja el informe de la **OCDE** (Ailbhe & OCDE, 2024). Paralelamente, iniciativas como **EMi** en Galicia (Servicio Público de Empleo Estatal, 2024) o **InfoSIA** (Asociación ASLAN, 2025) del Ayuntamiento de Madrid y el informe de la OCDE mencionado reflejan tendencias convergentes: un enfoque inicial en la elaboración de perfiles y orientación, un creciente interés en el emparejamiento (aunque con desafíos persistentes), la exploración de tecnologías más avanzadas como LLM y GenAI, así como un reconocimiento creciente de la importancia de factores como la explicabilidad, la participación y la gestión del cambio.

Es preciso aclarar que los casos citados anteriormente son ejemplos de iniciativas que, precisamente por las limitaciones metodológicas señaladas, no fueron identificados durante el proceso de revisión sistemática. Estos casos no forman parte del corpus de evidencia analizado, el cual mantiene su solidez y representatividad por los motivos ya expuestos. 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.

Estos estudios también ponen de manifiesto la importancia de factores como la explicabilidad, la participación de las partes interesadas y la necesidad de una planificación que considere la transición a la producción y la integración en los flujos de trabajo más amplios de los SPE.

> Un ejemplo concreto de aplicación en un SPE regional español (SPEG) es el motor de inteligencia laboral "EMi", que utiliza Procesamiento de Lenguaje Natural (PNL), ML y algoritmia predictiva para mejorar el emparejamiento entre la oferta y la demanda, analizar perfiles competenciales y estimar la empleabilidad.

Nuestra revisión, que partió de una lista definida de palabras clave y realizó búsquedas booleanas estructuradas, asegura la replicabilidad; cierto es que ésta podría no ser 'suficiente' si el objetivo es captar la totalidad del conocimiento relevante: **un enfoque iterando revisión desestructurada general o narrativa, entrevistas y, finalmente, la scoping, dotarían de más especificidad a la última**

A la luz de estas reflexiones, y considerando las limitaciones identificadas, se perfilan con claridad **posibles caminos de mejora** metodológica para futuros estudios en este ámbito.

Hubiera sido, por ejemplo, muy enriquecedor realizar fases preliminares antes de la scoping review formal: (i) una **revisión narrativa inicial**, menos estructurada, podría haber ayudado a identificar conceptos clave, debates principales y un espectro más amplio de terminología; (ii) una **consulta con expertos clave** (académicos, gestores públicos, técnicos implementando IA en servicios sociales/empleo) es fundamental para extraer conocimiento tácito y palabras clave a usar en la posterior revisión de alcance; (iii) y finalmente, la **revisión de alcance,** o scoping review propiamente dicha.

Ilustración 13 - Propuestas de mejora y lecciones aprendidas

Fuente: elaboración propia.

6. Bibliografía y referencias

6.1. Bibliografía

Ailbhe, B. & OCDE. (2024). *A new dawn for Public Employment Services. Service Delivery in the age of Artificial Intelligence* (No. 19; OCDE Artificial Intelligence Papers). https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/a-new-dawn-for-public-employment-services_25e1e70e/5dc3eb8e-en.pdf

Asociación ASLAN. (2025, February 11). InfoSIA: Inteligencia Artificial para la asistencia a profesionales en la propuesta de prestaciones y recursos idóneos a las personas usuarias de Servicios Sociales XVII Convocatoria Premios AAPP. *Asociación @aslan*. https://aslan.es/mejora-de-la-eficiencia-y-experiencia-del-empleado-candidatura2025/

Bajgar, M., & Criscuolo, C. (2019). Designing Evaluation of Modern Apprenticeships in Scotland. In N. Crato & P. Paruolo (Eds.), *Data-Driven Policy Impact Evaluation* (pp. 289–311). Springer International Publishing. https://doi.org/10.1007/978-3-319-78461-8_18

Better Evaluation. (2014). *Rainbow Framework*. https://www.betterevaluation.org/frameworks-guides/rainbow-framework

Bohni Nielsen, S., Mazzeo Rinaldi, F., & Petersson, G. J. (2024). *Artificial Intelligence and Evaluation: Emerging Technologies and Their Implications for Evaluation* (1st ed.). Routledge. https://doi.org/10.4324/9781003512493

Buttow, C. V. (2024). Data-Driven Policy Making and Its Impacts on Regulation: A Study of the OECD Vision in the Light of Data Critical Studies. *European Journal of Risk Regulation*, 1–19. https://doi.org/10.1017/err.2024.73

Crato, N., & Paruolo, P. (Eds.). (2019a). *Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design*. Springer International Publishing. https://doi.org/10.1007/978-3-319-78461-8

Crato, N., & Paruolo, P. (2019b). The Power of Microdata: An Introduction. In N. Crato & P. Paruolo (Eds.), *Data-Driven Policy Impact Evaluation* (pp. 1–14). Springer International Publishing. https://doi.org/10.1007/978-3-319-78461-8_1

d'Hombres, B., & Santangelo, G. (2019). Use of Administrative Data for Counterfactual Impact Evaluation of Active Labour Market Policies in Europe: Country and Time Comparisons. In N. Crato & P. Paruolo (Eds.), *Data-Driven Policy Impact Evaluation* (pp. 271–287). Springer International Publishing. https://doi.org/10.1007/978-3-319-78461-8_17

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Elevati, C. (2025). *73—L'Intelligenza Artificiale come alleato strategico per i professionisti MEAL |*. LinkedIn. https://www.linkedin.com/pulse/73-lintelligenza-artificiale-come-alleato-per-i-meal-elevati-4xvff/?trackingId=rvljDuu5T5yjRNJAMx9XPQ%3D%3D

European Commission. (2024). *Evaluation Handbook (2024)*. https://capacity4dev.europa.eu/library/evaluation-handbook-2024_en

European Commission. Directorate General for Employment, Social Affairs and Inclusion & ICF. (2025). *Opportunities of AI within PES processes and services: Exploring PES experiences, best practices and emerging business value.* Publications Office. https://data.europa.eu/doi/10.2767/84293

Harris, M., & Dawson, R. (2025, March). *Humans + AI: Macro Workflow Roles* [Red Social]. LinkedIn. https://www.linkedin.com/feed/update/urn:li:activity:7304825289916706816/

Ligero Lasa, J. Andrés. (2015). *Tres métodos de evaluación de programas y servicios*. Fundación Caja Madrid.

Martins, P. S. (2019). Public Policy, Big Data, and Counterfactual Evaluation: An Illustration from an Employment Activation Programme. In N. Crato & P. Paruolo (Eds.), *Data-Driven Policy Impact Evaluation* (pp. 149–163). Springer International Publishing. https://doi.org/10.1007/978-3-319-78461-8_10

Minguijon, J., & Serrano-Martinez, C. (2022). La Inteligencia Artificial en los Servicios Sociales: Estado de la cuestión y posibles desarrollos futuros. *Cuadernos de Trabajo Social*, *35*(2), Article 2. https://doi.org/10.5209/cuts.78747

Morescalchi, A., Van Veldhuizen, S., Voogt, B., & Vogt, B. (2019). Negative Home Equity and Job Mobility. In N. Crato & P. Paruolo (Eds.), *Data-Driven Policy Impact Evaluation* (pp. 183–202). Springer International Publishing. https://doi.org/10.1007/978-3-319-78461-8_12

Rodríguez Cano, N. S. (2018). Tendencias actuales en la evaluación de políticas públicas. *Ensayos de Economía*, *28*(53), 15–35. https://doi.org/10.15446/ede.v28n53.75382

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*. 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/laiaaplicadaalosserviciospublicosdeempleohaciaunsistemadeintermediacionorientacionyformacionbasadoencompetenciasprofesionales

Subirats, J. (Ed.). (2008). *Análisis y gestión de políticas públicas* (1. ed). Ariel.

Tangi, L., Van, N. C., Combetto, M., Gattwinkel, D., & Pignatelli, F. (2022). *AI Watch. European landscape on the use of Artificial Intelligence by the Public Sector* (JRC129301 No. EUR 31088 EN , OP KJ-NA-31088-EN-N (online)). European Union. https://doi.org/10.2760/39336

Züger, T., & Asghari, H. (2023). AI for the public. How public interest theory shifts the discourse on AI. *AI & SOCIETY*, *38*(2), 815–828. https://doi.org/10.1007/s00146-022-01480-5

6.2. Referencias incluídas en la muestra de la revisión

Achterhold, E., Mühlböck, M., Steiber, N., & Kern, C. (2025). Fairness in Algorithmic Profiling: The AMAS Case. *Minds and Machines*, *35*(1), 9. https://doi.org/10.1007/s11023-024-09706-9

Alexander, W. (2022). *Applying Artificial Intelligence to Public Sector Decision Making* [Major Research Paper]. University of Ottawa.

Arguelles Toache, E. (2023). Ventajas y desventajas del uso de la Inteligencia Artificial en el ciclo de las políticas públicas: Análisis de casos internacionales. *Acta universitaria*, *33*. https://www.redalyc.org/journal/416/41677664054/html/

Babšek, M., Ravšelj, D., Umek, L., & Aristovnik, A. (2025). Artificial Intelligence Adoption in Public Administration: An Overview of Top-Cited Articles and Practical Applications. *AI*, *6*(3), Article 3. https://doi.org/10.3390/ai6030044

Bamberger, M., & York, P. (2020). *Transforming Evaluation in the 4th Industrial Revolution: Exciting Opportunities and New Challenges* (pp. 11–21) [eVALUation Matters, Second Quarter 2020, 11-21.]. African Development Bank Group. https://idev.afdb.org/sites/default/files/documents/files/EM%20Q2-2020-article1-challenges%20and%20opportunities%204th%20industrial%20revolution%28En%29.pdf

Bilbao-Goyoaga, E. (2023). *Perceptions Matter: Quasi-Experimental Evidence on the Effects of Spain's New Minimum Income on Households' Financial Wellbeing* (No. Social Policy Working Paper 02-23; LSE Department of Social Policy.). https://www.lse.ac.uk/social-policy/Assets/Documents/PDF/working-paper-series/WPS-02-23-Eugenia-Bilbao-Goyoaga.pdf

Bouyousfi, S. E., & Ouedraogo, M. (2024). Artificial intelligence and big data-driven evaluation research and practices: A systematic literature review. *Evaluation*, 13563890241289937. https://doi.org/10.1177/13563890241289937

Carlizzi, D. N., & Quattrone, A. (2023). Artificial Intelligence and Data Governance for Precision ePolicy Cycle. In D. Marino & M. Monaca (Eds.), *Artificial Intelligence and Economics: The Key to the Future* (pp. 67–84). Springer International Publishing. https://doi.org/10.1007/978-3-031-14605-3_6

Cearns, J., & Knox, H. (2024). *Safeguarding Data*. https://doi.org/10.3167/cja.2024.420103

Dalla Longa, F., Sweerts, B., & van der Zwaan, B. (2021). Exploring the complex origins of energy poverty in The Netherlands with machine learning. *Energy Policy*, *156*, 112373. https://doi.org/10.1016/j.enpol.2021.112373

European Commission Joint Research Centre. (2020). *AI Watch—Artificial Intelligence in public services*. https://ai-watch.ec.europa.eu/publications/ai-watch-artificial-intelligence-public-services_en

European Commission. Joint Research Centre. (2024). *Competencies and governance practices for AI in the public sector.* Publications Office. https://data.europa.eu/doi/10.2760/7895569

Ferretti, S. (2023). Hacking by the prompt: Innovative ways to utilize ChatGPT for evaluators. *New Directions for Evaluation*, *2023*(178–179), 73–84. https://doi.org/10.1002/ev.20557

Franzen, S., Quang, C., Schweizer, L., Budzier, A., Gold, J., Vellez, M., Ramirez, S., & Raimondo, E. (2022). *Advanced Content Analysis: Can Artificial Intelligence Accelerate Theory-Driven Complex Program Evaluation?* (Independent Evaluation Group) [IEG Methods and Evaluation Capacity Development Working Paper Series]. World Bank.

Fundació iSocial. Innovació en l'acció social. (2024, septiembre 20). La Inteligencia Artificial en los Servicios Sociales: Análisis predictiva e identificación de necesidades de intervención. https://isocial.cat/es/la-inteligencia-artificial-en-los-servicios-sociales-analisis-predictiva-e-identificacion-de-necesidades-de-intervencion/

Gabrikova, B., Svabova, L., & Kramarova, K. (2023). Machine Learning Ensemble Modelling for Predicting Unemployment Duration. *Applied Sciences*, *13*(18), Article 18. https://doi.org/10.3390/app131810146

Germundsson, N., Stranz, H., & Bergmark, Å. (2024). Reducing administration? Examining the alignment of robotic process automation and social assistance in Swedish Personal Social Services. *Nordic Social Work Research*, 1–14. https://doi.org/10.1080/2156857X.2024.2440720

Hall, O., Dompae, F., Wahab, I., & 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*, *35*(7), 1753–1768. https://doi.org/10.1002/jid.3751

Hasan Chy, M. K., & Nana Buadi, O. (2024). Role of Machine Learning in Policy Making and Evaluation. *International Journal of Innovative Science and Research Technology (IJISRT)*, 456–463. https://doi.org/10.38124/ijisrt/IJISRT24OCT687

Head, C. B., Jasper, P., McConnachie, M., Raftree, L., & Higdon, G. (2023). Large language model applications for evaluation: Opportunities and ethical implications. *New Directions for Evaluation*, *2023*(178–179), 33–46. https://doi.org/10.1002/ev.20556

Hutter, R., & Hutter, M. (2021). Chances and Risks of Artificial Intelligence—A Concept of Developing and Exploiting Machine Intelligence for Future Societies. *Applied System Innovation*, *4*(2), Article 2. https://doi.org/10.3390/asi4020037

Jacob, S. (2025). Artificial Intelligence and the Future of Evaluation: From Augmented to Automated Evaluation. *Digit. Gov.: Res. Pract.*, *6*(1), 10:1-10:10. https://doi.org/10.1145/3696009

Kates, A. W., & Wilson, K. (2023). AI for Evaluators: Opportunities and Risks. *Journal of MultiDisciplinary Evaluation*, *19*(45), 99–104. https://doi.org/10.56645/jmde.v19i45.907

Kempeneer, S., Ranchordas, S., & 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. & ILO Regional Office for Asia and the Pacific. (2023). *Effects of digitalization on the human centricity of social security administration and services*. ILO. https://doi.org/10.54394/PMPD3825

Lehtiniemi, T. (2024). Contextual social valences for artificial intelligence: Anticipation that matters in social work. *Information, Communication & Society*, *27*(6), 1110–1125. https://doi.org/10.1080/1369118X.2023.2234987

m4social. (2024). Radar de algoritmos de IA y procesos de decisión automatizada para el acceso a los derechos sociales en Cataluña. *m4Social*. https://m4social.org/es/resources/radar-dalgoritmes-dia-i-processos-de-decisio-automatitzada-per-a-lacces-als-drets-socials-a-catalunya/

MacArthur, J., Moung, V., Carrard, N., & Willetts, J. (2025). Personas for program evaluation: Insights from a gender-focused evaluation in Cambodia. *Evaluation*, *31*(1), 70–91. https://doi.org/10.1177/13563890241284425

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*, *58*(5), 831–855. https://doi.org/10.1111/spol.12991

Mason, S. (2023). Finding a safe zone in the highlands: Exploring evaluator competencies in the world of AI. *New Directions for Evaluation*, *2023*(178–179), 11–22. https://doi.org/10.1002/ev.20561

OCDE. (2024). *Using AI to manage minimum income benefits and unemployment assistance: Opportunities, risks and possible policy directions* (OECD Artificial Intelligence Papers No. 21; OECD Artificial Intelligence Papers, Vol. 21). https://doi.org/10.1787/718c93a1-en

Perron, B. E., Hiltz, B. S., Khang, E. M., & Savas, S. A. (2025). AI-Enhanced Social Work: Developing and Evaluating Retrieval-Augmented Generation (RAG) Support Systems. *Journal of Social Work Education*, 1–11. https://doi.org/10.1080/10437797.2024.2411172

Perron, B. E., Luan, H., Victor, B. G., Hiltz-Perron, O., & 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*, 10497315241280686. https://doi.org/10.1177/10497315241280686

Picciotto, R. (2020). Evaluation and the Big Data Challenge. *American Journal of Evaluation*, *41*(2), 166–181. https://doi.org/10.1177/1098214019850334

Powell, S., Copestake, J., & Remnant, F. (2024). Causal mapping for evaluators. *Evaluation*, *30*(1), 100–119. https://doi.org/10.1177/13563890231196601

Rachovitsa, A., & 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*, *22*(2), ngac010. https://doi.org/10.1093/hrlr/ngac010

Ranerup, A., & Svensson, L. (2024). Value positions in the implementation of automated decision-making in social assistance. *Nordic Social Work Research*, *14*(4), 471–485. https://doi.org/10.1080/2156857X.2022.2062040

Ratner, H. F., & Schrøder, I. (2024). Ethical Plateaus in Danish Child Protection Services: The Rise and Demise of Algorithmic Models. *Science & Technology Studies*, *37*(3), Article 3. https://doi.org/10.23987/sts.126011

Ratner, H. F., & Thylstrup, N. B. (2024). Citizens' data afterlives: Practices of dataset inclusion in machine learning for public welfare. *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-01920-4

Raupach, M., Enderling, M., Feddoul, L., Legner, H., Mauch, M., & König-Ries, B. (n.d.). *Towards a semantic format for FIM: supporting German public services using the GerPS-FIM-Microverse ontology pipeline*.

Raveh, E., Ofek, Y., Bekkerman, R., & Cohen, H. (2020). Applying Big Data visualization to detect trends in 30 years of performance reports. *Evaluation*, 26(4), 516–540. https://doi.org/10.1177/1356389020905322

Ruggia-Frick, R. (2021). Applying emerging data-driven technologies in social security. Country experiences and ISSA guidelines. *Ubezpieczenia Społeczne. Teoria i Praktyka*, *150*(3). https://doi.org/10.5604/01.3001.0015.5230

Saura, J. R., & Debasa, F. (Eds.). (2022). *Handbook of Research on Artificial Intelligence in Government Practices and Processes:* IGI Global. https://doi.org/10.4018/978-1-7998-9609-8

Schimpf, C., Barbrook-Johnson, P., & Castellani, B. (2021). Cased-based modelling and scenario simulation for ex-post evaluation. *Evaluation*, *27*(1), 116–137. https://doi.org/10.1177/1356389020978490

Schmager, S., Grøder, C. H., Parmiggiani, E., Pappas, I., & Vassilakopoulou, P. (2024). Exploring citizens' stances on AI in public services: A social contract perspective. *Data & Policy*, 6, e19. https://doi.org/10.1017/dap.2024.13

Stern, E. (2020). Editorial. *Evaluation*, 26(4), 401–403. https://doi.org/10.1177/1356389020966442

Tilton, Z., Lavelle, J. M., Ford, T., & Montenegro, M. (2023). Artificial intelligence and the future of evaluation education: Possibilities and prototypes. *New Directions for Evaluation*. https://doi.org/10.1002/EV.20564

Väänänen, N. (2021). The digital transition of social security in Finland. Frontrunner experiencing headwinds? *Ubezpieczenia Społeczne. Teoria i Praktyka*, 4, 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*, *11*(1).

Valle-Cruz, D., Criado, J. I., Sandoval-Almazán, R., & Ruvalcaba-Gomez, E. A. (2020). Assessing the public policy-cycle framework in the age of artificial intelligence: From agenda-setting to policy evaluation. *Government Information Quarterly*, *37*(4), 101509. https://doi.org/10.1016/j.giq.2020.101509

York, P. (2024). The Future of Evaluation Analytics. In S. Bohni Nielsen, F. Mazzeo Rinaldi, & G. J. Petersson, *Artificial Intelligence and Evaluation* (1st ed., pp. 219–241). Routledge. https://doi.org/10.4324/9781003512493-11

York, P., & Bamberger, M. (2020). *Measuring Results and Impact in the Age of Big Data: The Nexus of Evaluation, Analytics, and Digital Technology*. The Rockefeller Foundation. https://www.rockefellerfoundation.org/reports/measuring-results-and-impact-in-the-age-of-big-data-the-nexus-of-evaluation-analytics-and-digital-technology/

Ziulu, V., Anuj, H., Hagh, A., Raimondo, E., & Vaessen, J. (2024). Extracting Meaning from Textual Data for Evaluation. In S. Bohni Nielsen, F. Mazzeo Rinaldi, & G. J. Petersson, *Artificial Intelligence and Evaluation* (1st ed., pp. 78–102). Routledge. https://doi.org/10.4324/9781003512493-5

7. Anexos metodológicos

7.1. Anexo I. Algoritmos booleanos de búsqueda

7.1.1. Implementación

7.1.1.1 Google académico:

("artificial intelligence" OR "AI" OR "machine learning" OR "ML") ("social policy" OR "welfare policy" OR "public administration" OR "government program" OR "social work*" OR "social service*") ("minimum income" OR "guaranteed income" OR "basic income" OR "income support" OR "renta garantizada") ("implementation" OR "case study" OR "application" OR "evaluation") -school -university -classroom -student -educational -education -medical -health -healthcare -hospital -tourism -marketing

Filtro de años: 2019- (actualidad)

Y también:

("artificial intelligence" OR "AI" OR "machine learning") ("minimum income" OR "guaranteed income" OR "IMV" OR "RGI") ("implementation" OR "case study") -education -health -marketing

Filtro de años: 2019- (actualidad)

7.1.1.2 SCOPUS

208 documents found

TITLE-ABS-KEY (("artificial intelligence" OR "AI" OR "ML" OR "machine learning") AND ("use case*" OR "application*" OR "implementation*") AND ("social polic*" OR "IMV" OR "guaranteed income" OR "minimum income*" OR "guarantee* minimum income" OR "employment" OR "welfare" OR "social inclusion" OR "basic income scheme*" OR "income support" OR "minimum income guarantee" OR "RGI" OR "Renta de Garantia de Ingresos" OR "prestacion* de garantia de ingreso*" OR "renta* minima*") AND NOT ("education" OR "health" OR "terrorism" OR "finance" OR "marketing" OR "medic*")) AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND (LIMIT-TO (AFFILCOUNTRY , "Italy") OR LIMIT-TO (AFFILCOUNTRY , "United Kingdom") OR LIMIT-TO (AFFILCOUNTRY , "Spain") OR LIMIT-TO (AFFILCOUNTRY , "Germany") OR LIMIT-TO (AFFILCOUNTRY , "Greece") OR LIMIT-TO (AFFILCOUNTRY , "France") OR LIMIT-TO (AFFILCOUNTRY , "Netherlands") OR LIMIT-TO (AFFILCOUNTRY , "Portugal") OR LIMIT-TO (AFFILCOUNTRY , "Sweden") OR LIMIT-TO (AFFILCOUNTRY , "Slovakia") OR LIMIT-TO (AFFILCOUNTRY , "Czech Republic") OR LIMIT-TO (AFFILCOUNTRY , "Austria") OR LIMIT-TO (AFFILCOUNTRY , "Slovenia") OR LIMIT-TO (AFFILCOUNTRY , "Hungary") OR LIMIT-TO (AFFILCOUNTRY , "Belgium") OR LIMIT-TO (AFFILCOUNTRY , "Ireland") OR LIMIT-TO (AFFILCOUNTRY , "Finland") OR LIMIT-TO (AFFILCOUNTRY , "Croatia") OR LIMIT-TO (AFFILCOUNTRY , "Romania") OR LIMIT-TO (AFFILCOUNTRY , "Lithuania") OR LIMIT-TO (AFFILCOUNTRY , "Bulgaria") OR LIMIT-TO (AFFILCOUNTRY , "Latvia") OR LIMIT-TO (AFFILCOUNTRY , "Denmark")

OR LIMIT-TO (AFFILCOUNTRY , "Malta") OR LIMIT-TO (AFFILCOUNTRY , "Cyprus") OR LIMIT-TO (AFFILCOUNTRY , "Switzerland")) AND (LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "DECI") OR LIMIT-TO (SUBJAREA , "MULT") OR LIMIT-TO (SUBJAREA , "PSYC") OR LIMIT-TO (SUBJAREA , "ENGI") OR LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "MATH") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "ECON") OR LIMIT-TO (SUBJAREA , "ARTS")) AND (LIMIT-TO (LANGUAGE , "English") OR LIMIT-TO (LANGUAGE , "Spanish")) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re") OR LIMIT-TO (DOCTYPE , "bk"))

WOS 372 results from All Databases for:

TS=(("artificial intelligence" OR "AI" OR "ML" OR "machine learning")

AND

("use case*" OR "application*" OR "implementation*")

AND

("social polic*" OR "IMV" OR "guaranteed income" OR "minimum income*" OR "guarantee* minimum income" OR "employment" OR "welfare" OR "social inclusion" OR "basic income scheme*" OR "income support" OR "minimum income guarantee" OR "RGI" OR "Renta de Garantia de Ingresos" OR "prestacion* de garantia de ingreso*" OR "renta* minima*")

NOT

("education" OR "health" OR "terrorism" OR "finance" OR "marketing" OR "medic*"))

7.1.1.3 Google

After:2020 ("USE CASE" OR "EXAMPLE" OR "IMPLEMENT*" OR "APPLICATION") AND ("ARTIFICIAL INTELLIGENCE" OR "MACHINE LEARNING" OR "ALGORITHM*") AND AND ("social service*" OR "SOCIAL INCLUSION" OR "SOCIAL POLICY" OR "SOCIAL POLICIES" OR "LABOUR INCLUSION" OR "LABOR INCLUSION")

After:2020 ("EVALUACION" OR "CASO DE USO" OR "EJEMPLO" OR "IMPLEMENT*" OR "USE CASE") AND (AI OR IA OR "INTELIGENCIA ARTIFICIAL" OR ML OR "MACHINE LEARNING" OR "ALGORITM*" OR "APRENDIZAJE AUTOMÁTICO") AND ("IMV" OR "INGRESO MÍNIMO VITAL" OR "RENTA* GARANTIZADA*" OR "GARANTIA* DE INGRESO*" OR "RGI" OR "Renta de Garantía de Ingresos" OR "prestacion* de garantía de ingreso*" OR "renta* mínima*") AND ("EMPLEO" OR "EMPLEABILIDAD" OR "INCLUSIÓN SOCIAL" OR "INCLUSIÓN LABORAL")

After:2020 ("CASO DE USO" OR "EJEMPLO" OR "IMPLEMENT*" OR "USE CASE") AND (AI OR IA OR "INTELIGENCIA ARTIFICIAL" OR ML OR "MACHINE LEARNING" OR "ALGORITM*" OR "APRENDIZAJE AUTOMÁTICO") AND ("INGRESO MÍNIMO VITAL" OR "RENTA* GARANTIZADA*" OR "GARANTIA* DE INGRESO*" OROR "Renta de Garantía de Ingresos" OR "prestacion* de garantía de ingreso*" OR "renta* mínima*") AND ("EMPLEO" OR "EMPLEABILIDAD" OR "INCLUSIÓN SOCIAL" OR "INCLUSIÓN LABORAL")

After:2020 ("USE CASE" OR "EXAMPLE" OR "IMPLEMENT*" OR "APPLICATION") AND ("ARTIFICIAL INTELLIGENCE" OR "MACHINE LEARNING" OR "ALGORITHM*") AND ("MINIMUM INCOME" OR "GUARANTEED INCOME" OR "BASIC INCOME" OR "UNIVERSAL BASIC INCOME" OR "MINIMUM INCOME GUARANTEE" OR "INCOME SUPPORT") AND ("EMPLOYMENT" OR "EMPLOYABILITY" OR "SOCIAL INCLUSION" OR "LABOUR INCLUSION" OR "LABOR INCLUSION")

After:2020 ("USE CASE" OR "EXAMPLE" OR "IMPLEMENT*" OR "APPLICATION") AND ("ARTIFICIAL INTELLIGENCE" OR "MACHINE LEARNING" OR "ALGORITHM*") AND AND ("social service*" OR "SOCIAL INCLUSION" OR "SOCIAL POLICY" OR "SOCIAL POLICIES" OR "LABOUR INCLUSION" OR "LABOR INCLUSION")

7.1.2. *Evaluación*

7.1.2.1 Google académico

evaluat*

AND ("public policy cycle" OR "public polic*")

AND ("artificial intelligence" OR "machine learning")

AND (“employment polic*” OR job* OR “social inclusion” OR “social polic*”)

158 resultados (0,26 s) --> 23 abstracts-título seleccionados

7.1.2.2 Google

("EVALUACION" OR "EVALUACIÓN" OR “EVALUA*” OR “IMPACTO*”)

AND ("INTELIGENCIA ARTIFICIAL" OR "APRENDIZAJE AUTOMÁTICO" OR "MACHINE LEARNING” OR “BIG DATA” OR “DATA SCIENCE” OR “CIENCIA DE DATOS”)

AND ("politica* publica*” OR “POLITICA* SOCIAL*” OR “CICLO DE LA POLÍTICA PÚBLICA” OR “CICLO DE LAS POLÍTICAS PÚBLICAS”)

AND ("EMPLEO" OR "EMPLEABILIDAD" OR "INCLUSIÓN SOCIAL" OR "INCLUSIÓN LABORAL")

80 resultados

AFTER:2019 ESPAÑA ("EVALUACION" OR "EVALUACIÓN" OR "EVALUA*" OR "IMPACTO*" OR "IMPLEMENTA*")

AND ("INTELIGENCIA ARTIFICIAL" OR "APRENDIZAJE AUTOMÁTICO" OR "MACHINE LEARNING" OR "BIG DATA" OR "DATA SCIENCE" OR "CIENCIA DE DATOS" OR "ALGORITM*")

AND ("politica* publica*" OR "POLITICA* SOCIAL*" OR "PROTECCIÓN SOCIAL" OR "SISTEMA* DE BIENESTAR")

AND ("EMPLEO" OR "EMPLEABILIDAD" OR "INCLUSIÓN SOCIAL" OR "INCLUSIÓN LABORAL" OR "MERCADO LABORAL")

AND ("INTELIGENCIA ARTIFICIAL" OR "APRENDIZAJE AUTOMÁTICO" OR "MACHINE LEARNING” OR “BIG DATA” OR “DATA SCIENCE” OR “CIENCIA DE DATOS”)

AND ("politica* publica*" OR "POLITICA* SOCIAL*" OR "CICLO DE LA POLÍTICA PÚBLICA" OR "CICLO DE LAS POLÍTICAS PÚBLICAS")

AND ("EMPLEO" OR "EMPLEABILIDAD" OR "INCLUSIÓN SOCIAL" OR "INCLUSIÓN LABORAL")

160resultados

After:2020

("EVALUACION" OR "EVALUACIÓN" OR "EVALUA*" OR "IMPACTO*")

AND ("INTELIGENCIA ARTIFICIAL" OR "APRENDIZAJE AUTOMÁTICO" OR "MACHINE LEARNING" OR "BIG DATA" OR "DATA SCIENCE" OR "CIENCIA DE DATOS")

AND ("politica* publica*" OR "POLITICA* SOCIAL*" OR "CICLO DE LA POLÍTICA PÚBLICA" OR "CICLO DE LAS POLÍTICAS PÚBLICAS")

AND ("EMPLEO" OR "EMPLEABILIDAD" OR "INCLUSIÓN SOCIAL" OR "INCLUSIÓN LABORAL")

30 resultados

7.1.2.3 Otras búsquedas

Snowball (Bohni Nielsen et al., 2024) y posteriormente en Litmaps (citas de las citas; aplicables a Evaluación). 55 resultados; 24 resultados relevantes recuperados.

7.1.2.3.1 SIIS

(I) "inteligencia artificial AND evaluación AND políticas sociales": 3 resultados, 2 priorizados (relevancia 3-4).

(II) "Inteligencia artificial AND evaluación": 15 resultados, 5 priorizados (relevancia 3-4).

(III) "Inteligencia artificial AND evaluación OR políticas públicas": 26 resultados, 5 priorizados.

7.1.2.3.2 EVALUATION

"Artificial intelligence": 9 resultados, 2 priorizados.

"Artificial intelligence" OR "machine learning": 18 resultados, 6 priorizados.

7.2. Anexo II. Resumen de hallazgos en el ámbito de la Implementación

Tabla 4 - Resumen de hallazgos en la primera revisión de implementación

Fuente (Año)	Contexto	Tema Central	Tipo de IA/Datos/Método Aplicado	Resultados Clave	Limitaciones técnicas, organizacionales y éticas de las tecnologías
Achterhold et al. (2025)	Austria (2018-2020). Sistema AMAS implementado por el Servicio Público de Empleo	Sesgos y equidad en el sistema algorítmico AMAS	Evaluación empírica de equidad y aplicación de técnicas de mitigación de sesgos	1) El sistema AMAS muestra rendimiento insuficiente (33%erróneas); 2) Existe sesgo de género; 3) Las técnicas de mitigación de sesgos redujeron efectivamente las disparidades	Limitación sociodemográfica del grupo (jóvenes); medidas de apoyo insuficientemente definibles; limitación severa en métricas de oportunidad y en equidad
Cearns & Knox (2024)	Etnografía ML/AI en gobierno local UK (austeridad)	Autoridad local UK ("Summertown", zona deprimida)	Sistema ML (proveedor privado) integra datos multiagencia para perfiles riesgo ("flags") que ayudan a trabajadores soc.	Consenso ("bien público"). Miedo a no usar datos > mal uso. Visiones distintas sobre datos. Modelos ven ciudadanos estáticos. Datos añaden contexto. Tensión: dato como recurso (ok) vs. decisor autónomo (mal).	Conflictos interpretación datos, motivos distintos (ayuntamiento/proveedor), preocupación ética (profiling), tensión humano vs. dato, definición "dato" varía.
Dalla Longa et al. (2021)	ML para predecir riesgo pobreza energética (NED)	Países Bajos (sin política específica vs. UK)	Algoritmo ML aprende ponderación factores socioeconómicos desde datos para predicción de riesgo.	Ingreso: predictor principal. Factores vivienda secundarios. Precisión ~80% (f1). Relaciones complejas entre factores.	Difícil identificar vulnerables subrepresentados, necesita datos grandes/representativos, mecanismos causales complejos, elección umbrales ML afecta rendimiento.
European Commission. JRC. (2024)	Gobernanza y competencias IA en sector público UE	Varios países/municipios UE (2017-23)	Diversas apps IA (email, chatbots, predicción crimen/precios/personal)	Éxitos (clasif. email ITA, hojas ruta, registros IA-AMS); Fracasos (riesgo infantil DIN, chatbots baja uso, adaptación comercial difícil)	Financiación, escasez talento, silos/preparación datos (80% trabajo), interpretación reglas, falta transparencia algoritmos.
European Commission Joint Research Centre (2020)	Gobernanza y competencias IA en sector público UE	Varios países/municipios UE (aprox.2012-2019)	230 casos (muchos, fuera del alcance temporal); algunos de ellos, en el ámbito de empleo e inclusión)	Enfoque en asistencia. Resultados de mayor eficiencia administrativa y optimización operativa	Retos en protección de privacidad, ética algorítmica, protección del trabajo y transparencia algorítimica
Fundació iSocial. Innovació en l'acció social (2024)	wSocial (Cataluña) y Nrativ (Suecia)	Prestación de servicios; detección de necesidades	2 casos, en EU, 2024. Detección de la vulnerabilidad (España) y mejora continua de servicio administrativo (Suecia)	Capacidad para etectar necesidades urgentes, optimizar la distribución de recursos y facilitar la intervención anticipada	No se señalan

Gabrikova et al., (2023)	Oficina Central de Trabajo, Asuntos Sociales y Familia de la República Eslovaca (COLSAF)	Predicción de la duración del desempleo	Modelos de aprendizaje automático combinados (ensemble): árboles de decisión (CART, CHAID), análisis discriminante Uso del método "stacking"	Predicción eficaz del desempleo (78% precisión general, 92% precisión casos largos	Caja negra. Necesidad de gobernanza de datos. Actualización de datos constante.
Germundsson et al., (2024)	Sistema de asistencia social sueco, donde cuatro municipios han implementado RPA en sus servicios personales sociales entre 2018 y 2020	Estudio cualitativo basado en entrevistas grupales con trabajadores sociales	Automatización de Procesos Robóticos (RPA) de nivel básico, que opera con reglas predefinidas e interactúa con interfaces digitales existentes.	La adopción de RPA requirió importantes ajustes organizativos y administrativos. Las mejoras de eficiencia percibidas se atribuyeron más a cambios organizativos que a la tecnología en sí	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 et al. (2024)	IA en serv. sociales/fraude; impacto desproporcionado en vulnerables	Varios países UE/USA; agencias gob./municipales	Ejemplos: Servicio proactivo (BEL), fraude (DIN), discriminación (NED escándalo), datos malos (UK), chatbot gradual (UK)	Simplificación/ayuda (BEL); Daño masivo (NED escándalo); "Imprudencia" al mantener sistemas defectuosos.	Mala calidad datos (ignorada), sesgo/racismo institucional (NED), eficiencia > equidad, incumplimiento RGPD (frecuente).
Lee-Archer & ILO (2023)	Interoperabilidad datos en seg. social (BEL)	Bélgica (sistema nacional seg. social; ~3000 actores)	Plataforma interoperabilidad datos segura (SOA, estándares abiertos)	Premio ONU 2006 (e-gobierno); mejora servicio ciudadanos/empleadores.	No indicadas en la fuente proporcionada.
Lehtiniemi, Tuukka (2024)	ML para predecir riesgo infantil/juvenil (FIN)	Finlandia (bienestar infantil; piloto 2021)	Herramienta ML predice riesgo custodia/emergencia usando cientos de "factores riesgo" (datos admin.)	Info vista como útil/objetiva por trab. sociales; filtra/resalta datos clave para atención.	Problemas razonamiento abductivo (asume presente=producto del pasado); modelos reducen personas a "conjunto de características".
m4social Radar Report (2024)	Mapeo IA/ADM en Adm. Cat. (DD.SS.); 12 algoritmos	Cataluña (2023-24); Adm. Cat. (BCN/Gen/AOC)	12 sist. IA: mayoría autom. admin., supervisados, auditados. Ej: DPR (demandas), RisCanvi (reincidencia), VideoID, Pobr.Energ.NLP, clasificación, recon. facial, chatbots, reglas, predicción.	+Calidad, +respuesta, +eficiencia, +accesibilidad, -tiempos/costes. Ej: DPR 75-95% concord.	Sist. internos (garantías); cautela (explicab.); necesidad de monitorizar y evaluar impacto real.
Marienfeldt, Justine (2024)	Impacto herramientas digitales en burócratas ventanilla	Países nórdicos + GER, NED, GRE	Herramientas digitales diversas (comunicación, gestión casos, evaluación riesgo, automatización)	Libera rutina pero despersonaliza servicio; transforma rol ("burócrata pantalla"); aumenta carga ciudadano/brecha digital.	Diseño rígido (ignora complejidad social/individualidad); aplicación inflexible de reglas universales.

Minguijón & Serrano-Martínez (2022)	Adaptación SS españoles a IA; modelo evaluación	Sistema SS España (multi-actor)	Modelo evalúa IA en fases intervención SS y grado automatización.	Éxitos: Georef. (VLC), Info (BCN), IA colectiva (BCN), Coord. (CyL), Hist. Social (AND), Planif. (BCN).	Necesario Data Warehouse; participación prof. Trabajo Social; implementación gradual; interoperabilidad clave; apoyo público decidido.
OCDE (2024)	Implementación IA en prestaciones sociales (OCDE)	Agencias gubernamentales UE (varios países)	Diversa: Reconocimiento voz (AUT), Chatbot (BEL, FIN), ML (DEN, GER), Reglas (EST), Algoritmos (NED), Patrones (UK).	Eficiencia (EST), alivio carga (BEL), detección proactiva (DEN, UK, NED), mejor asignación recursos (AUT, CZE), asistencia (FIN), optimización docs (GER).	Sesgo automatización, privacidad, transparencia (solo NED con registro explícito), definición IA (EST no es IA estricta), evaluación impacto limitada.
Perron, Brian E. et al. (2024)	Rendimiento LocalLLM en NLP para invest. trabajo social	Invest. trabajo social (bienestar infantil, datos sensibles)	4 LocalLLMs (Mistral, Mixtral, Llama3) vía Ollama, comparados con codificadores humanos expertos.	Llama3 destaca; rendimiento similar a humano (más rápido); pocas alucinaciones (~1%); clasificación > extracción.	Prompts no sistemáticos (posible sesgo Llama3), obsolescencia rápida resultados, errores en "verdad campo", restricciones LLM propietario para temas sensibles.
Perron, Brian E. et al. (2025)	RAG para mejorar asesoramiento académico	Asesoramiento universitario (acceso info compleja/dispersa)	RAG: permite consultas lenguaje natural (resúmenes) + control directo contenido por expertos. Base: manual estudiante.	Alta precisión (>91%), sin evidencia alucinaciones/sesgo; errores menores del LLM considerados insignificantes en la práctica.	Evaluación limitada (1 caso), ventana contexto LLM, necesita más testeo pre-producción, trade-off flexibilidad con plataformas no-code.
Rachovitsa & Johann (2022)	Implicaciones DDHH de algoritmos (Caso SyRI, NED)	Gobierno/municipios Países Bajos (barrios específicos)	SyRI: Cruce masivo datos personales para perfiles de riesgo de fraude (modelo opaco).	Tribunal: SyRI violó privacidad (Art. 8 CEDH), falto de transparencia y salvaguardas, riesgo discriminatorio.	Opacidad deliberada, sin notificación a afectados, riesgo discriminación (targeting), datos amplios, obstaculiza revisión judicial.
Ranerup & Svensson (2024)	ADM en asistencia social (municipios Suecia)	4 municipios Suecia (2018-20)	Infraest. TIC (e-solicitud, RPA, gestión casos, seguimiento) + Reorganización roles trabajadores sociales.	+Eficacia recursos, -costos; discreción afectada por infraestructura digital; +rendición de cuentas.	Metodología (no análisis x grupo, intenciones), diferentes valores profesionales, problemas técnicos/tiempo.
Ratner & Schrøder (2024)	ML en sector público (perfiles algorítmicos, DIN)	Dinamarca (registros exhaustivos)	Sistema predicción riesgo maltrato infantil (ML sobre datos históricos administrativos).	No claramente reportados	Restricciones legales fusión datos; riesgo sobreinformación sin base legal; necesidad refinar datos vs. sesgos identificados (ej. étnico).
Raupach et al. (n.d.)	Digitalización integral procesos admin. (OZG, Alemania)	Alemania (estructura federal compleja)	Kit desarrollo (métodos/software) para ecosistema digitalización FIM; UI web; representación semántica servicios.	Beneficios potenciales reportados: modernizar registros, simular implicaciones legales, detectar vacíos/costos, feedback rápido pre-ley.	No claramente reportados

Saura & Debasa (2022)	Adopción IA en gobierno digital; modelos adopción	Ejemplos UE (BEL, FIN, FRA, UK)	30 casos: NLP (CitizenLab), personalización (AuroraAI), ciudad inteligente, robot, control voz, teledetección ML.	Clave: Explicabilidad. Baja preocupación por privacidad/seguridad/sesgo. Alta beneficencia en servicios. Adopción: intención, facilidad, valor percibido. Confianza gobierno influye. Uso regulatorio: alto impacto social.	Dependencia datos secundarios, posible sesgo publicación (éxitos), generalización limitada.
Schmager, S. et al. (2024)	IA responsable en serv. públicos (ética, contrato social)	Principalmente Noruega (+AUT, NED)	Prototipo IA ("lentes contrato social") predice duración baja médica como apoyo informativo para trabajadores sociales.	Actitud general (+); prototipo mejoró actitudes (+); preocupación por automatización total; valor IA para eficiencia; IA aún opaca para algunos participantes.	Barreras IA (rendimiento, gobernanza datos), técnicas/gestión, regulatorias; riesgo sesgo; "caja negra"; reuso datos; difícil adaptación contexto público; implementación real limitada.
Väänänen (2021)	Digitalización seg. social Finlandia (foco IA/ADM en Kela)	Finlandia (Kela, alta confianza admin., estado bienestar nórdico)	Arquitectura digital nacional interconectada; Kela (RPA, IA convers.), otros (ML predicción pensión discapacidad); muchas decis. auto.	Alta digitalización (72% solicitudes online); ML predice 78% pensiones discapacidad; registros centralizados buenos para ML.	Legalidad RPA/ADM cuestionada (Defensor Pueblo: "decisiones no solo automáticas"); vacío legal vs. Constitución; nueva legislación necesaria/urgente según instituciones.
Valle Escolano (2023)	Ambivalencia IA vs. pers. discapacidad (beneficios/riesgos)	Marco global/europeo (normativa)	Análisis sist. IA / decisión algorítmica en ámbitos discapacidad (empleo, educ., salud, vida indep.). ADM, recon. facial/emocional, ML, clasificación/predicción.	(+): Apoyo, movilidad, comunicación, inclusión, accesibilidad. (-): Discriminación algorítmica, opacidad, exclusión, capacitismo.	Equipos diversos; transparencia; involucrar usuarios; gobernanza ética (DDHH); regulación específica; combatir sesgo; investigar inclusión; IA como innovación social.

7.3. Anexo III. Resumen de hallazgos en el ámbito de la evaluación

Tabla 5 Resumen de hallazgos en casos de uso para evaluación (casos de uso).

Autor(es) y Año	Contexto	Tema Central	Tipo de IA/Datos/Método Aplicado	Resultados Clave	Limitaciones técnicas, organizacionales y éticas de las tecnologías
Alexander (2022)	Administración pública canadiense	Análisis del impacto de la inteligencia artificial (IA) en los procesos de toma de decisiones del sector público	sistemas de toma de decisiones automatizados (ADM) potenciados por IA en el sector público	Aumento significativo de la eficiencia organizacional; se prevé una transformación organizacional fundamental	Limitaciones éticas; problemas de explicabilidad, sesgo algorítmico y disminución de la rendición de cuentas
Arguelles Toache (2023)	Colaboración interinstitucional entre organismos académicos, centros de investigación tecnológica y entidades multilaterales de cooperación al desarrollo	Mapas de pobreza	Machine Learning aplicado al reconocimiento de patrones en datos heterogéneos, visión artificial para procesamiento de imágenes satelitales, y algoritmos de correlación	Mayor frecuencia de actualización de datos comparado con censos quinquenales o decenales y reducción significativa de costos de recogida de datos. Capacidad de monitoreo cuasi-continuo	Sesgos muestrales, validación limitada, eficacia diferencial según contexto
Babšek et al. (2025)	Revisión de la literatura y bibliometría	Análisis del estado actual de la adopción de inteligencia artificial (IA) en la administración pública	Se analizan múltiples subconjuntos de IA	La investigación muestra un aumento significativo en la producción científica sobre IA en administración pública desde 2018, con China y EE.UU. liderando. El aprendizaje automático es el subconjunto de IA más estudiado	Desarrollar marcos de gobernanza ética que prioricen transparencia, equidad y rendición de cuentas en sistemas de IA; Implementar evaluaciones de impacto obligatorias para algoritmos antes de su despliegue; Equilibrar innovación tecnológica con cumplimiento regulatorio
Bamberger & York (2020)	Global/África (2019-20)	Transformación de la evaluación por la 4ª Rev. Industrial (IA, Big Data).	Discusión sobre ML (cuasi-experimentos, automatización) y NLP (análisis cualitativo).	Oportunidades: datos rápidos/diversos, análisis de poblaciones completas, complejidad, sistemas, sostenibilidad. NLP 8x más rápido que humanos.	Falta de familiaridad de evaluadores, costes, reorganización, resistencia, brecha digital, ética/privacidad de datos.

Bilbao-Goyoaga (2023)	España (2020-22)	Impacto del Ingreso Mínimo Vital (IMV).	ML Supervisado (Control Sintético).	Corto plazo: sin efecto económico real. Medio plazo: mejora percepción subjetiva. Posible efecto positivo posterior (pendiente).	Alto non take-up (57%), reducción de presupuestos autonómicos, retraso y limitación de datos, posible efecto placebo.
Bouyousfi & Ouedraogo (2024)	Revisión sistemática de 102 artículos primarios seleccionados de un total de 227 identificados.	Análisis global de literatura académica sobre IA y big data en evaluación, abarcando desde enero 2004 hasta agosto 2023	Varios: modelos de lenguaje grandes (LLMs) como GPT, aprendizaje automático (supervisado y no supervisado), redes neuronales, análisis de sentimiento, etc.	Falta de representatividad en los datos utilizado no reflejan la población general; dificultades para establecer causalidad mediante correlaciones en big data; opacidad de algoritmos y modelos complejos	Sistemas integrados, combinación de ciencia de datos con enfoques tradicionales, utilizando IA y big data como investigación exploratoria complementada con métodos convencionales
Carlizzi & Quattrone (2023)	Ámbito global con énfasis en la administración pública europea; publicado en 2023, con particular atención a las iniciativas de la UE (EU4FACTS y Policy Cloud).	Análisis del impacto de la Inteligencia Artificial (IA) en la transformación del ciclo de políticas públicas tradicional, proponiendo un nuevo modelo denominado "Precision ePolicy Cycle"	Varios: análisis y categorización masiva de datos, sistemas de soporte a la decisión, aprendizaje automático (ML) y aprendizaje profundo (DL), IA explainable (XAI), gemelos digitales, procesamiento de lenguaje natural, análisis de sentimiento,	El proceso sigue a la vida social a través de datos en lugar de ser un modelo abstracto y lineal; reduce a tres las fases "tradicionales". Las políticas se vuelven adaptativas, personalizadas y sensibles a los datos	Riesgos de manipulación de datos; comprensión limitada de la IA; cajas negras; necesidad de readaptación de los procesos y las organizaciones; riesgo de tecnocracia; perfilados ilegítimos.
Ferretti (2023)	Global (2022-23)	Uso de ChatGPT en evaluación.	LLM (ChatGPT-3/4).	Usos potenciales: generar docs, diseñar instrumentos, análisis texto, comunicación, democratización.	Alucinaciones, falta de verificación factual, sesgos en entrenamiento, riesgo de sustituir juicio crítico, privacidad/ética, falta de control.
Franzen et al. (2022)	Banco Mundial (64 países)	IA para análisis de contenido en evaluación compleja (desnutrición).	ML Supervisado (clasif. texto), ML No Supervisado (topic modeling), t-SNE, Grafos Conocimiento.	ML sup.: precisión modesta (etiquetas exactas), alta (categorías). ML no sup.: excelente identificación temas clave. Grafos: estructuran output.	Inversión tiempo inicial, extracción manual de texto, datos desequilibrados, complejidad grafos, evidencia incompleta.
Hasan Chy & Nana Buadi (2024)	Entorno de políticas públicas cada vez más complejo	Análisis del potencial transformador del machine learning (ML) para mejorar los procesos de formulación y	El artículo examina cómo el ML puede transformar los procesos de formulación y evaluación de políticas públicas en	Capacidad del ML para automatizar análisis, proporcionar capacidades predictivas, similar escenarios, analizar sentimiento público, etc.	Problemas de calidad de datos, sesgo algorítmico, falta de interpretabilidad

		evaluación de políticas públicas	sectores como salud, economía, medio ambiente, educación, justicia criminal y seguridad web.		
Head et al. (2023)	Análisis teórico y revisión conceptual con estimaciones de impacto basadas en opiniones de expertos	Aplicaciones de los modelos de lenguaje grande (LLMs) en evaluación: oportunidades, limitaciones técnicas e implicaciones éticas	Análisis de modelos de lenguaje grande (LLMs) como ChatGPT y sus aplicaciones en tareas evaluativas	Las tareas más impactadas serían análisis y redacción (46% de reducción) y diseño de evaluación (40% de reducción)	Limitaciones técnicas: inexactitud factual ("alucinaciones"), pobre rendimiento numérico, opacidad del procesamiento (problema de "caja negra"). Problemas éticos: perpetuación de sesgos, exclusión de lenguas minoritarias, costo medioambiental
Jacob (2025)	Análisis teórico-conceptual y prospectivo	Evolución del uso de la inteligencia artificial en la evaluación de políticas públicas: desde la evaluación aumentada hasta la potencial automatización	Varios: IA generativa (como ChatGPT), procesamiento de lenguaje natural (NLP), etc.	Identificación de aplicaciones actuales de IA en todas las fases del proceso de evaluación (planificación, recolección y análisis de datos, formulación de juicios y recomendaciones, informes y difusión).	Desafíos técnicos: problema de "caja negra", falta de transparencia y explicabilidad en los sistemas de IA, alucinaciones en IA generativa, necesidad de verificación humana. Problemas éticos: posible refuerzo de desigualdades sociales y económicas, sesgos algorítmicos, protección de datos personales.
MacArthur et al. (2024)	Camboya (saneamiento)	"Personas evaluativas" para análisis cualitativo.	NLP (LDA, STM) para validar personas creadas manualmente.	Creación de 14 personas; ayudó a visualizar datos complejos, facilitó discusiones, dio anonimato parcial.	Posible embellecimiento en narrativas retrospectivas; problemas de anonimato con muestreo censal (mitigado parcialmente).
Mason (2023)	Campo profesional de la evaluación en el contexto de la creciente adopción de tecnologías de IA generativa	Análisis conceptual y desarrollo de marcos teóricos para clasificar competencias evaluativas según su vulnerabilidad ante la IA	Análisis conceptual sobre diferentes tipos de IA (inteligencia estrecha, amplia y general) y sus implicaciones para el campo de la evaluación	Las competencias interpersonales (Dominio 5.0) y contextuales (Dominio 3.0) tienen menor riesgo de ser reemplazadas por IA,	Reconocimiento de la "Paradoja de Moravec": la dificultad para predecir qué tareas serán fáciles o difíciles de automatizar para la IA, con sorpresas constantes en el desarrollo tecnológico
Picciotto (2020)	Campo profesional de la evaluación y el análisis de datos masivo	Ensayo crítico y análisis conceptual sobre las oportunidades y desafíos de la integración del Big Data en la evaluación	Análisis de algoritmos predictivos y sistemas de Big Data aplicados en diversos sectores (justicia criminal, servicios sociales,	Crítica al enfoque de correlación sin teoría ("el fin de la teoría"); problemas de representatividad y exclusión digital; dilemas éticos de privacidad; tensión entre precisión y equidad en algoritmos	Identificación de múltiples riesgos asociados con el uso no regulado del Big Data: toma de decisiones automatizada injusta, algoritmos sesgados, modelos defectuosos,

			educación, gestión laboral)		manipulación de la opinión pública, y prácticas laborales opresivas.
Powell et al. (2024)	Campo profesional de la evaluación y el análisis de datos masivo	Presentación del mapeo causal como herramienta metodológica para capturar, codificar y visualizar afirmaciones causales interconectadas en evaluaciones	Mención al procesamiento de lenguaje natural (NLP) como herramienta complementaria	Desarrollo de un marco metodológico de mapeo causal adaptado específicamente para evaluadores	Ambigüedad en las etiquetas de los factores causales; riesgo de "trampa de transitividad" al inferir relaciones causales indirectas de distintos contextos; dificultad para capturar sistemáticamente la fuerza o tipo de influencia causal
Raveh et al. (2020)	Israel (1987-2017)	Detección automática de tendencias en informes de desempeño.	NLP, ML No Supervisado (clasif.), ESA, reducción dimensionalidad, visualización grafos.	Identificó clusters temáticos y tendencias sin keywords predefinidas; redujo significativamente la revisión manual.	Metodología innovadora (sin hipótesis previas) es la principal aportación; no se detallan limitaciones específicas del resultado.
Schimpf et al. (2021)	Académico (análisis 2008-18)	Modelado basado en casos y simulación (CBSS) para evaluación ex-post.	ML No Supervisado (k-means, SOMs), Simulación Monte Carlo.	CBSS modela complejidad, identifica subgrupos, explora contrafactuales, evalúa robustez, complementa eval. realista.	Limitaciones inherentes a modelos, no causalidad definitiva, requiere buena especificación y datos, depende de interacción evaluador.
Stern (2020)	Académico,profesional de la evalauación	Editorial; académico	Mención breve de herramientas automatizadas para el análisis de datos no estructurados	No aplica	Medición de resultados intangibles en el sector cultural; las limitaciones del análisis costo-beneficio en evaluaciones de impacto; y los desafíos de las evaluaciones basadas en probabilidades e inferencias causales
Tilton et al. (2023)	Educación Eval. (2022-23)	IA (chatbots) en la formación de evaluadores.	LLM (GPT-3.5 turbo), NLP, RAG.	Prototipos informaron, sintetizaron y adaptaron respuestas sobre evaluación. Rendimiento > expectativas iniciales.	Técnicas (traducción, contexto, alucinaciones), Metodológicas (eval. preliminar, sin métricas), Éticas (sesgos, equidad acceso).
Valle-Cruz et al.(2020)	Revisión. Implementación global de tecnologías de inteligencia artificial en gobiernos y administraciones públicas,	Análisis del impacto de la inteligencia artificial en las diferentes etapas del ciclo de políticas públicas y propuesta de un modelo dinámico alternativo al tradicional	Propuesta de un modelo de Ciclo Dinámico de Políticas Públicas (DPPC) que incorpora la IA en forma de espiral incremental	Mayor precisión, eficiencia y velocidad en la definición de agenda; mejora del procesamiento de información y toma de decisiones en la formulación; automatización inteligente en la implementación; y monitoreo en tiempo real en la evaluación	Opacidad, brecha digital y exclusión; sesgos algorítmicos potenciales; falta de transparencia; violación de privacidad; complejidad de los algoritmos; dependencia excesiva de la IA

York (2024)	EE. UU., centrado en la región de Pittsburgh y en servicios de salud mental infantil	Aplicación de Modelado Causal Estructural (SCM) y Analítica de Precisión (PA), para mejorar la evaluación de programas sociales. A través de dos estudios de caso	Structural Causal Modeling (SCM) y Precision Analytics (PA); Machine Learning supervisado	Buenos resultados clínicos; Identificación de combinaciones de intervenciones más efectivas para subgrupos específicos;	Falsos positivos/negativos. Dependencia de la calidad y exhaustividad de los datos disponibles. Limitaciones técnicas
York & Bamberger (2020)	África y el contexto global de desarrollo;2019-2020.	Algoritmos de aprendizaje automático en evaluación	Varias: 1) Big data, 2) Internet de las Cosas (IoT), 3) Drones y satélites, 4) Inteligencia artificial y minería de datos, y 5) Blockchain.	Acceso a datos más rápido, económico y diverso; posibilidad de trabajar con universo en lugar de muestras;Incorporación de variables contextuales para análisis de complejidad; Integración de análisis de sistemas	Costes e inversión; retos organizacionales y familiaridad con las tecnologías; brecha digital; propiedad y uso de los datos
Ziulu et al., (2024)	Evaluación temática de los indicadores relativos a los años 2010-20 del programa de préstamos Doing Business (DB) del Banco Mundial.	El Independent Evaluation Group (IEG) desarrolló una metodología híbrida que combina NLP supervisado, búsquedas por palabras clave y revisión manual selectiva	Minería de Texto y Machine Learning Supervisado	El enfoque permitió duplicar el tamaño de la base de evidencia identificada (269 proyectos en total), mejorando así la validez y amplitud de los hallazgos. Reducción del tiempo; reutilización de modelos	Las limitaciones pueden afectar a la calidad de los datos; dificultades en captura completa de datos textuales; necesidad de equipos interdisciplinares

besaldi

Enplegu eta inklusio politiken ebaluazio organoa

Órgano de evaluación de las políticas de empleo e inclusión

EUSKO JAURLARITZA
GOBIERNO VASCO

EKONOMIA, LAN ETA ENPLEGU SAILA

DEPARTAMENTO DE ECONOMÍA, TRABAJO Y EMPLEO