

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

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**

Casos de uso

Elaborado con la asistencia técnica de



Tabla de contenido

0.	INTRODUCCIÓN	3
1.	SISTEMA DE PERFILADO PARA ESTÍMULOS AL EMPLEO AMAS (AUSTRIA).....	4
2.	IA PARA SERVICIOS SOCIALES, DE INCLUSIÓN Y EMPLEO (CATALUÑA).....	5
3.	SISTEMA PREDICTIVO DE BAJAS POR ENFERMEDAD (NORUEGA).....	8
4.	DIGITALIZACIÓN Y USO DE LA IA EN LOS SERVICIOS PÚBLICOS (FINLANDIA).....	9
5.	MODELOS DE LENGUAJE GRANDE (LLM) PARA EXTRAER DATOS SENSIBLES (EE. UU.)	10
6.	EVALUACIÓN DE RIESGOS, APOYO INFORMATIVO Y ATENCIÓN CIUDADANA (PLURINACIONAL).....	11
7.	MODELO PREDICTIVO DE RIESGO DE POBREZA ENERGÉTICA (PAÍSES BAJOS).....	13
8.	MODELOS PREDICTIVOS DE DURACIÓN DEL DESEMPLEO (ESLOVAQUIA).....	14
9.	IMÁGENES SATELITALES PARA ANÁLISIS DE POBREZA (PLURINACIONAL)	15
10.	ANALÍTICA DE PRECISIÓN EN EVALUACIÓN DE PROGRAMAS SOCIALES (EE. UU.)	16
11.	MINERÍA DE TEXTO Y MACHINE LEARNING PARA EVALUACIÓN TEMÁTICA (PLURINACIONAL).....	17
12.	INGENIERÍA DE DATOS Y MACHINE LEARNING PARA MEDIR IMPACTO EN POBREZA (PLURINACIONAL).....	18
13.	AUTOMATIZACIÓN DE INFORMES DE VULNERABILIDAD ENERGÉTICA (CATALUÑA)	19
14.	DECISIONES AUTOMATIZADAS (ADM) EN EL SECTOR PÚBLICO (CANADÁ).....	20
15.	EXPERIMENTOS NATURALES CON DATOS HISTÓRICOS (PLURINACIONAL).....	21
16.	MACHINE LEARNING PARA EVALUAR EL IMPACTO DEL IMV (ESPAÑA)	22
17.	MODELO DE LENGUAJE GRANDE (LLM) EN EVALUACIÓN (ESTUDIO ANALÍTICO)	23
18.	MACHINE LEARNING PARA UNA EVALUACIÓN BASADA EN LA TEORÍA DEL CAMBIO (PLURINACIONAL).....	24
19.	PERSONAS EVALUATIVAS: NLP Y MODELOS DE APRENDIZAJE AUTOMÁTICO (CAMBOYA).25	
20.	BIG DATA PARA DETECTAR TENDENCIAS EN DESEMPEÑO (ISRAEL).....	26
21.	CASE BASED MODELLING (CBM) Y SIMULACIÓN DE ESCENARIOS (REINO UNIDO).....	27
22.	CHATBOTS PARA LA FORMACIÓN DE EVALUADORES (EE. UU. Y CANADÁ)	28

o. Introducción

Los casos de uso ilustrados en este documento complementan el informe final del **estudio exploratorio sobre buenas prácticas, innovaciones y transformaciones en la aplicación de la Inteligencia Artificial (IA) en el ciclo de la política pública**.

Este documento recopila **22 fichas ilustrativas de casos** proporcionando ejemplos concretos y detallados de cómo la IA se ha aplicado en diferentes contextos y fases del ciclo de las políticas públicas de empleo e inclusión social. Las fichas ilustran, respectivamente, 9 casos de aplicación en la fase de implementación de la política y 13 en la fase de evaluación.

Los campos desglosados en cada ficha incluyen:

1. **Caso de uso:** describe la aplicación específica de la IA, proporcionando una visión general del sistema o tecnología implementada. Este campo es crucial para entender el propósito y la funcionalidad de la IA en el contexto de la política pública.
2. **Contexto:** proporciona información sobre el entorno en el que se implementa la IA, incluyendo detalles sobre la organización responsable y el marco temporal. Este campo ayuda a situar la aplicación de la IA en un contexto real y práctico.
3. **Descripción general:** ofrece una explicación detallada del funcionamiento del sistema de IA, incluyendo los procesos y métodos utilizados. Este campo es esencial para comprender cómo se implementa la tecnología y qué problemas específicos aborda.
4. **Tecnología aplicada:** detalla las tecnologías y algoritmos específicos utilizados en la aplicación de la IA. Este campo es útil para los profesionales que buscan entender las bases técnicas de la implementación y las herramientas utilizadas.
5. **Resultados documentados:** presenta los resultados y logros obtenidos a través de la implementación de la IA, incluyendo métricas de éxito y beneficios observados. Este campo es fundamental para evaluar la efectividad y el impacto de la tecnología en el contexto de la política pública.
6. **Riesgos y limitaciones:** identifica los posibles desafíos, riesgos y limitaciones asociados con la implementación de la IA. Este campo es importante para una evaluación crítica y equilibrada de la tecnología, destacando áreas que requieren atención y mejora.
7. **Referencias:** incluye la fuente bibliográfica de la publicación que ilustra el caso de uso.

Este compendio de 22 casos prácticos permite al lector comprender cómo la Inteligencia Artificial puede transformar las políticas de empleo e inclusión social, mostrando tanto beneficios como riesgos, y ofreciendo una visión práctica para quienes planean incorporar estas herramientas en la administración pública.

1. Sistema de perfilado para estímulos al empleo AMAS (Austria)

Caso de uso	Sistema algorítmico para ayudar a los trabajadores sociales a decidir qué solicitantes de empleo deberían recibir recursos de apoyo. El sistema calcula una puntuación de "Oportunidad de Integración" para cada persona, clasificándolas en tres grupos según su probabilidad de reintegrarse al mercado laboral. Esto determina qué nivel de apoyo reciben: a más alta probabilidad de reempleo, menos recursos.
Contexto	Sistema implementado por el Servicio Público de Empleo de Austria (AMS) en 2018, que buscaba optimizar tanto la efectividad de los programas de empleo como la eficiencia administrativa.
Descripción general	Se descubrió que el modelo penalizaba sistemáticamente a mujeres y personas con ciudadanía no europea: al combinar datos administrativos con características personales para predecir probabilidades de reempleo, el sistema reforzó desigualdades estructurales bajo una apariencia de objetividad técnica. El sistema fue suspendido en 2021 por decisión judicial. El estudio ilustra los sesgos y las posibilidades para su mitigación.
Tecnología aplicada	Algoritmo basado en regresión logística.
Resultados documentados	<ul style="list-style-type: none"> ▪ Técnicas de mitigación. El estudio ilustra diversas técnicas de mitigación de sesgos (reponderación, representaciones justas, post-procesamiento de oportunidades equalizadas), y demuestra que estos métodos pueden reducir las disparidades entre grupos mientras mantienen niveles aceptables de rendimiento predictivo. ▪ Métricas de equidad. El estudio ilustra métricas de igualdad como medida para identificar y mitigar el sesgo algorítmico, entre ellas, Statistical Parity Difference (SPD) y Disparate Impact (DI), Equal Opportunity Difference (EOD) y Average Odds Difference (AOD).
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ El estudio analiza una versión reconstruida del algoritmo AMAS, no el original, y se limita al género como atributo binario sin considerar otras características protegidas ni perspectivas interseccionales. ▪ Su metodología se basa en métricas observacionales que ignoran relaciones causales, priorizando la equidad grupal sobre la individual. ▪ Utiliza datos históricos potencialmente sesgados por políticas anteriores y emplea técnicas de mitigación poco transparentes que no permiten identificar qué elementos específicos son modificados durante el proceso.
Referencias	Achterhold, E., Mühlböck, M., Steiber, N., & Kern, C. (2025). Fairness in Algorithmic Profiling: The AMAS Case. <i>Minds and Machines</i> , 35(1), 9. https://doi.org/10.1007/s11023-024-09706-9

2. IA para servicios sociales, de inclusión y empleo (Cataluña)

Ayuntamiento de Barcelona, Área de Derechos Sociales e Instituto Municipal de Servicios Sociales. Sistema DPR (Identificación automática de demandas, problemas y respuestas sociales).

Caso de uso	Automatización del registro de datos en intervenciones sociales mediante procesamiento de lenguaje natural.
Contexto	Sistema de apoyo a decisiones basado en inteligencia artificial avanzada, implementado en 40 centros de servicios sociales que atienden aproximadamente 50.000 primeras visitas anuales.
Descripción general	Utiliza procesamiento del lenguaje natural para agilizar el registro de datos en el sistema de información. Tras grabar conversaciones con usuarios, el sistema identifica automáticamente demandas (D), problemas (P) y recursos (R) para optimizar el tiempo de los profesionales.
Tecnología aplicada	Procesamiento de lenguaje natural aplicado a datos no estructurados (transcripciones de entrevistas sociales).
Resultados documentados	<ul style="list-style-type: none">▪ Implementación exitosa desde octubre 2021 para todos los centros de Barcelona, servicio de atención a la dependencia y Equipo de Demandas de Estudio de Infancia y Adolescencia en Riesgo. En 2022 se internalizó en la infraestructura corporativa del Ayuntamiento.▪ Se recomienda el establecimiento de procesos de medición de calidad y reentrenamiento tanto automático como manual que permiten monitorizar el éxito del algoritmo.
Riesgos y limitaciones	<ul style="list-style-type: none">▪ Posibilidad de sesgos.▪ Modelo algorítmico alojado en un servidor externo.

Generalitat de Catalunya, Departamento de Derechos Sociales. Motor de reglas de prestaciones sociales

Caso de uso	Verificación automatizada de cumplimiento de criterios para prestaciones sociales.
Contexto	Administración pública catalana, sistema implementado y en uso.
Enfoque	Sistema basado en reglas para evaluar elegibilidad.
Descripción general	Verifica si un ciudadano o entidad cumple los criterios establecidos para recibir prestaciones sociales específicas. Incluye cinco algoritmos desarrollados para: necesidades básicas, gastos del hogar, PNC jubilación, complemento de la PNC jubilación y prestación por nacimiento, acogida y adopción.
Tecnología aplicada	Algoritmos basados en reglas.
Resultados documentados	<ul style="list-style-type: none">▪ Servicios públicos (centrados en el ciudadano: mayor accesibilidad del ciudadano a los servicios públicos)▪ Capacidad de respuesta de la operación gubernamental

		<ul style="list-style-type: none"> Mejora de la gestión de los recursos públicos Aumento de la calidad de procesos y sistemas
Riesgos y limitaciones		No reportados.

Generalitat de Catalunya, Departamento de Derechos Sociales. Identificación necesidades en la intervención social

Caso de uso	Identificación automática de necesidades durante intervenciones sociales.
Contexto	Sistema en fase piloto integrado en la herramienta eSocial.
Enfoque	Asistencia en tiempo real para trabajadores sociales.
Descripción general	Durante la intervención social, el empleado público puede activar (con consentimiento del ciudadano) una funcionalidad que transcribe la conversación e identifica automáticamente situaciones de necesidad.
Tecnología aplicada	Procesamiento de lenguaje natural y posiblemente análisis predictivo.
Resultados documentados	No reportados (sistema en fase piloto).
Riesgos y limitaciones	No reportados.

Administración Abierta de Cataluña (AOC). Xatbots conversacionales.

Caso de uso	Asistentes virtuales para atención ciudadana y soporte a empleados públicos.
Contexto	Administración Abierta de Cataluña (AOC), sistema implementado y en uso para atención 24/7.
Enfoque	Automatización de asistencia y resolución de consultas.
Descripción general	Asistentes conversacionales desarrollados para proporcionar atención continua a empleados públicos y ciudadanía. Responden preguntas e incidencias relacionadas con trámites digitales y uso del portal de transparencia.
Tecnología aplicada	Algoritmos de IA conversacional entrenados con datos existentes de preguntas frecuentes.
Resultados documentados	<ul style="list-style-type: none"> Optimización de recursos: los chatbots reducen la carga de trabajo al gestionar consultas frecuentes; disminuyen gastos en atención presencial o telefónica. Mayor eficiencia: responden de forma inmediata y precisa a múltiples usuarios. Análisis de datos: permiten detectar patrones y necesidades para mejorar servicios. Acceso 24/7: información y asistencia disponibles en todo momento.

	<ul style="list-style-type: none"> ▪ Respuesta inmediata: sin esperas, agiliza trámites y consultas. ▪ Autorresolución: los usuarios resuelven dudas sin intervención humana. ▪ Mejora de experiencia: navegación guiada y respuestas precisas facilitan la interacción con la Administración.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Dependencia tecnológica: fallos o interrupciones pueden afectar la atención y generar desconfianza. ▪ Privacidad y seguridad: existe riesgo de vulneración de datos, especialmente por el registro de conversaciones. ▪ Limitaciones de asistencia: dificultades en la comprensión del lenguaje o en responder adecuadamente afectan la calidad del servicio. ▪ Calidad de respuestas: sin buen entrenamiento, los chatbots pueden dar respuestas incorrectas, afectando la confianza y eficacia del sistema.

La publicación incluye 12 sistemas de IA utilizados por cuatro administraciones públicas catalanas. De estos, seis operan a nivel local y seis a nivel autonómico; siete están implementados y cinco en fase piloto. La mayoría (7 de 12) buscan automatizar tareas administrativas. Los sistemas identificados abarcan desde identificación automática de demandas sociales hasta verificación de elegibilidad para prestaciones y asistentes conversacionales.

Se utilizan diferentes tipos de IA y sistemas tecnológicos de análisis de datos avanzados, incluyendo procesamiento de lenguaje natural, sistemas basados en reglas, asistentes conversacionales y algoritmos de recomendación. Destaca que 11 de los 12 sistemas mantienen algún tipo de supervisión humana.

La referencia para todos los casos detallados es: 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-laccess-als-drets-socials-a-catalunya/>.

3. Sistema predictivo de bajas por enfermedad (Noruega)

Caso de uso	Prototipo interactivo que simula un portal de servicios públicos con secuencias de interacción predefinidas para estimar la duración de bajas por enfermedad.
Contexto	Una organización pública noruega que gestiona distintos tipos de prestaciones sociales adopta un sistema predictivo con enfoque Human-in-the-loop (HITL), donde el sistema de IA actúa como fuente complementaria de información para trabajadores sociales, no como sistema de decisión automatizada.
Descripción general	El estudio describe la experiencia piloto y busca explorar el impacto de la adopción de IA en los servicios públicos, abordando tanto los beneficios como los desafíos éticos y regulatorios derivados de su uso, y cómo la transparencia y el involucramiento humano (human-in-the-loop) puede mejorar su aceptación social. Participan en el piloto 20 ciudadanos de entre 18 y 65 años, seleccionados para reflejar la distribución demográfica de la población en situación de baja por enfermedad.
Tecnología aplicada	El prototipo se basa en un sistema predictivo integrado en un portal interactivo.
Resultados documentados	<ul style="list-style-type: none"> Actitud mayoritariamente positiva hacia la implementación de IA en servicios públicos. Factores contribuyentes: tranquilidad por la supervisión humana y por la transparencia de procesos (nivel micro); propuesta de valor equilibrada entre beneficios individuales, como evitar citas innecesarias, y colectivos, en términos de una mejor gestión de recursos públicos (nivel meso); alta confianza preexistente en los órganos gubernamentales; equipo dedicado a explorar posibilidades de análisis de datos e IA; información clara sobre usos de datos y lógica del modelo. Implicaciones para el diseño de sistemas: necesidad de articular claramente la propuesta de valor tanto para individuos como para la sociedad. Supervisión humana: relevancia de mantener "humanos en el circuito" para preservar la legitimidad percibida.
Riesgos y limitaciones	<ul style="list-style-type: none"> Riesgo de pérdida de confianza si las respuestas no son claras o adecuadas. Preocupación por la privacidad y uso indebido de datos personales. Dificultad para comprender explicaciones técnicas o legales del sistema. Posible dependencia excesiva en la tecnología, reduciendo la atención personalizada.
Referencia	Schmager, S., Grøder, C. H., Parmiggiani, E., Pappas, I., & Vassilakopoulou, P. (2024). Exploring citizens' stances on AI in public services: A social contract perspective. <i>Data & Policy</i> , 6, e19. https://doi.org/10.1017/dap.2024.13

4. Digitalización y uso de la IA en los servicios públicos (Finlandia)

Caso de uso	Integración de tecnologías avanzadas en procesos administrativos en los servicios públicos de: seguridad social y salud.
Contexto	Finlandia posee una larga tradición en el uso intensivo de tecnología en su administración pública, respaldada por altos niveles de confianza ciudadana y un modelo de bienestar social universal, facilitando una integración efectiva de tecnologías avanzadas en procesos administrativos.
Descripción general	El análisis muestra cómo Finlandia ejemplifica una integración efectiva de tecnologías digitales avanzadas en servicios públicos, subrayando la importancia crítica de un marco regulatorio claro para garantizar transparencia, privacidad y confianza pública.
Tecnología aplicada	<ul style="list-style-type: none"> ▪ Inteligencia Artificial (IA) para predicción y personalización. ▪ Machine learning para evaluación de riesgo (p. ej., pensión por discapacidad). ▪ Automatización de procesos mediante Robotic Process Automation (RPA). ▪ Chatbots inteligentes para atención al cliente. ▪ Arquitectura nacional digital (AuroraAI) para servicios interconectados. ▪ Blockchain para verificación descentralizada de datos. ▪ Plataformas digitales integradas (Suomi.fi, Kanta.fi, Tulorekisteri.fi).
Resultados documentados	<ul style="list-style-type: none"> ▪ Incremento significativo en eficiencia y personalización del servicio público, destacando la aceptación positiva de los ciudadanos (72% solicitudes en línea en 2020). ▪ Implementación exitosa de IA para predecir riesgos de discapacidad laboral, con precisión del 78%. ▪ Desarrollo continuo hacia la automatización total, incluyendo la previsión y manejo eficiente de casos rutinarios mediante RPA.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Desafíos legales relacionados con decisiones totalmente automatizadas, requiriendo legislación específica. ▪ Problemas éticos y de privacidad relacionados con la transparencia y el uso extensivo de datos personales. ▪ Persistencia del uso de documentos físicos y resistencia parcial a la completa digitalización por parte de algunos ciudadanos.
Referencia	Väänänen, N. (2021). The digital transition of social security in Finland. Frontrunner experiencing headwinds? <i>Ubezpieczenia Społeczne. Teoria i Praktyka</i> , 4, 71–86. https://doi.org/10.5604/01.3001.0015.5251

5. Modelos de Lenguaje Grande (LLM) para extraer datos sensibles (EE. UU.)

Caso de uso	Aplicación de modelos de lenguaje locales (LocalLLMs) para clasificar y extraer información sobre problemas relacionados con abuso de sustancias (SRPs) a partir de resúmenes no estructurados de investigaciones en casos de abuso y negligencia infantil. El objetivo de la aplicación es mejorar el análisis de grandes volúmenes de texto confidencial sin comprometer la privacidad.
Contexto	El estudio se realiza en Estados Unidos, utilizando informes de una agencia estatal de bienestar infantil.
Descripción general	El estudio examina cuatro LocalLLM y mide el rendimiento de los modelos en tareas de clasificación y extracción textual, comparándolos con la codificación realizada por expertos humanos en 2.956 resúmenes previamente codificados manualmente.
Tecnología aplicada	LocalLLMs: modelos de lenguaje grande ejecutables en ordenadores locales, con técnicas que permiten realizar tareas sin entrenamiento específico previo.
Resultados documentados	<ul style="list-style-type: none">▪ Los LocalLLMs alcanzaron resultados comparables a los expertos humanos, con una exactitud del 95% y un kappa de Cohen de 0.90.▪ Los modelos mostraron alta fidelidad (99-100%) para casos de verdaderos positivos y tasas generales de alucinación bajas (0.6-2.9%). Incluso los modelos más pequeños (Llama3-8b) demostraron rendimiento excepcional, permitiendo su ejecución en hardware de nivel consumidor.▪ Los modelos exhibieron alta adherencia a las instrucciones (96-100%) para la generación de citas y sustancias en formatos estructurados.▪ Alta relevancia en la información extraída, facilitando la interpretación correcta de los casos.
Riesgos y limitaciones	<ul style="list-style-type: none">▪ Proceso iterativo no sistemático para el desarrollo y modificación de prompts.▪ Requisitos de hardware especializado (GPU).▪ LocalLLMs generan texto basado en patrones aprendidos, malentendidos sobre la generación de texto vs. extracción pura.▪ Brecha entre la formación tradicional en trabajo social y las habilidades técnicas necesarias para implementar estas tecnologías.▪ Falsos positivos relacionados con medicamentos no necesariamente indicativos de problemas de sustancias.
Referencia	Perron, B. E., et al. (2024). Moving Beyond ChatGPT: Local Large Language Models (LLMs) and the Secure Analysis of Confidential Unstructured Text Data in Social Work Research. <i>Research on Social Work Practice</i> , 10497315241280686. https://doi.org/10.1177/10497315241280686

6. Evaluación de riesgos, apoyo informativo y atención ciudadana (plurinacional)

Caso de uso	Integración de herramientas digitales y de IA para el apoyo en la toma de decisiones en los servicios públicos de: empleo, servicios sociales, atención ciudadana, control migratorio, regulación fiscal, educación.
Contexto	Se han analizado casos de: Dinamarca, Alemania, Grecia, Países Bajos, Noruega, Suecia.
Descripción general	Se trata de una revisión sistemática que analiza cómo la introducción de herramientas de IA integrada en un contexto general de análisis de datos avanzados transforma la toma de decisiones de trabajadores públicos, considerando tanto restricciones como potenciales mejoras en discrecionalidad y calidad del servicio, así como los cambios en la interacción entre ciudadanos y administración.
Tecnología aplicada	<ul style="list-style-type: none"> ▪ Sistemas automatizados para toma de decisiones. ▪ Herramientas de evaluación de riesgos. ▪ Sistemas digitales de gestión de casos. ▪ Bases de datos estructuradas. ▪ Herramientas de autoservicio ciudadano.
Resultados documentados	<ul style="list-style-type: none"> ▪ Automatización libera tiempo para tareas complejas. ▪ Reducción significativa de la interacción presencial directa. ▪ Incremento en la eficiencia administrativa y reducción de errores humanos. ▪ Riesgo de estandarización excesiva y pérdida de flexibilidad profesional.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Resistencia y estrategias informales por parte de funcionarios para afrontar rigidez tecnológica. ▪ Crecimiento en brecha digital y desigualdad en acceso ciudadano a servicios digitales. ▪ Riesgos éticos asociados a decisiones automatizadas sin adecuada supervisión humana.
Referencia	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. <i>Social Policy & Administration</i> , 58(5), 831–855. https://doi.org/10.1111/spol.12991

La publicación analiza casos de distintos contextos geográficos:

- ▶ Dinamarca: herramientas de evaluación de riesgos, base de datos y gestión de casos en el ámbito de los servicios de empleo; archivo digital en el ámbito de los servicios sociales.
- ▶ Alemania: sistema de gestión del conocimiento y de casos en el ámbito de los servicios de empleo y la atención ciudadana.
- ▶ Grecia: base de datos con formularios en el ámbito de los centros de atención ciudadana.
- ▶ Países Bajos: sistema inteligente (cámaras) y evaluación de riesgos en el ámbito del control migratorio.
- ▶ Noruega: autoservicio digital y gestión de casos en el ámbito de los servicios de empleo y de regulación fiscal.
- ▶ Suecia: decisiones automatizadas, evaluación estandarizada y gestión de casos en el ámbito de la asistencia social, becas y empleo.

La información incluida en la ficha se refiere al análisis conjunta de casos.

7. Modelo predictivo de riesgo de pobreza energética (Países Bajos)

Caso de uso	El estudio analiza la aplicación de diferentes tecnologías de aprendizaje automático en la predicción del riesgo de pobreza energética: un fenómeno de alta complejidad y multicausal.
Contexto	Los Países Bajos carecen de mecanismos oficiales para monitorear, cuantificar y enfrentar la pobreza energética, lo que motiva el desarrollo de un modelo predictivo basado en datos socioeconómicos.
Descripción general	El estudio destaca cómo los modelos de Machine Learning pueden contribuir significativamente a la comprensión y gestión de la pobreza energética, aunque subraya la necesidad de políticas claras y un enfoque más detallado en la selección de las variables clave.
Tecnología aplicada	<ul style="list-style-type: none"> ▪ Machine Learning mediante XGBoost. ▪ Clasificación algorítmica basada en datos socioeconómicos (ingreso, valor de la vivienda, tamaño del hogar, etc.). ▪ Uso de bases de datos socioeconómicas (KWB y WoON).
Resultados documentados	<ul style="list-style-type: none"> ▪ Alta capacidad predictiva del modelo (alrededor del 80% de precisión). ▪ Identificación clara de factores clave como ingresos, valor y antigüedad de la vivienda, tamaño del hogar y densidad poblacional. ▪ Confirmación de que los ingresos son el factor más determinante, aunque otros factores tienen una significativa capacidad predictiva.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Sensibilidad del modelo a la selección y exclusión de variables específicas. ▪ Necesidad de grandes volúmenes de datos representativos para entrenar adecuadamente los modelos, especialmente para identificar grupos subrepresentados. ▪ Complejidad intrínseca en la relación entre pobreza energética y factores socioeconómicos que requiere análisis más profundos.
Referencia	Dalla Longa, F., Sweerts, B., & van der Zwaan, B. (2021). Exploring the complex origins of energy poverty in The Netherlands with machine learning. <i>Energy Policy</i> , 156, 112373. https://doi.org/10.1016/j.enpol.2021.112373

8. Modelos predictivos de duración del desempleo (Eslovaquia)

Caso de uso	Predicción de la duración del desempleo para clasificar a los demandantes de empleo según su riesgo de desempleo de larga duración.
Contexto	Oficina Central de Trabajo, Asuntos Sociales y Familia de la República Eslovaca (COLSAF), responsable de gestionar el registro y apoyo a personas desempleadas.
Descripción general	El estudio desarrolla un modelo de aprendizaje automático que predice la duración del desempleo en Eslovaquia usando datos reales de demandantes de empleo. Se aplica un enfoque de modelos en ensamblado (CART, CHAID y análisis discriminante) combinados mediante regresión logística. El objetivo es apoyar políticas públicas y optimizar recursos para reducir el desempleo de larga duración.
Tecnología aplicada	<ul style="list-style-type: none"> Modelos de aprendizaje automático combinados (ensemble): árboles de decisión (CART, CHAID), análisis discriminante. Uso del método "stacking" (combinar resultados de varios modelos para mejorar la precisión).
Resultados documentados	<ul style="list-style-type: none"> Predicción eficaz del desempleo (78% precisión general, 92% precisión casos largos). Factores clave identificados: historial laboral de desempleo, educación y edad. Potencial para mejorar las intervenciones laborales mediante predicciones certeras.
Riesgos y limitaciones	<ul style="list-style-type: none"> Algunos modelos predictivos son poco transparentes o difíciles de interpretar ("caja negra"). Necesidad de actualización continua de datos para mantener precisión.
Referencia	Gabrikova, B., Svabova, L., & Kramarova, K. (2023). Machine Learning Ensemble Modelling for Predicting Unemployment Duration. <i>Applied Sciences</i> , 13(18), 10146. https://doi.org/10.3390/app131810146

9. Imágenes satelitales para análisis de pobreza (plurinacional)

Caso de uso	Predicción de pobreza o niveles de bienestar mediante el análisis de imágenes satelitales, aplicando algoritmos de IA para identificar indicadores como infraestructura, calidad de vivienda, iluminación nocturna o uso de móviles.
Contexto	Estudios internacionales con aplicaciones en países de África, Asia y América Latina. Aunque no se centra en un único país, se incluyen casos de Senegal, Bangladesh, Ruanda, Ghana, Uganda, etc. Los modelos tienen aplicación potencial en agencias de desarrollo, planificación y estadística.
Descripción general	El estudio analiza 60 trabajos que combinan imágenes satelitales e inteligencia artificial para predecir pobreza. Muestra que la combinación de modelos ML y DL, el uso de indicadores físicos y múltiples fuentes de datos mejora la precisión. Ofrece pautas para su aplicación en políticas públicas de desarrollo.
Tecnología aplicada	<ul style="list-style-type: none"> ▪ Aprendizaje automático (Machine Learning) y profundo (Deep Learning), redes neuronales. ▪ Extracción y selección de características mediante técnicas avanzadas (CNN, Random Forests, PCA). ▪ Imágenes satelitales para obtener información sobre pobreza.
Resultados documentados	<ul style="list-style-type: none"> ▪ Mejores predicciones de pobreza al utilizar indicadores físicos visibles (infraestructura, calidad de vivienda, etc.) frente a indicadores sociales variables (ingreso, consumo). ▪ Limitado efecto positivo al aumentar la resolución espacial de las imágenes. ▪ Potencial de mejorar significativamente la rapidez y precisión de evaluaciones de pobreza.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Baja interpretabilidad de algunos métodos avanzados (necesidad de IA explicable). ▪ Mayor precisión al combinar técnicas tradicionales y avanzadas, indicando necesidad de métodos mixtos.
Referencia	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. <i>Journal of International Development</i> , 35(7), 1753-1768. https://doi.org/10.1002/jid.3751

10. Analítica de precisión en evaluación de programas sociales (EE. UU.)

Caso de uso	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: (1) Gemma Services, personalización de tratamientos en salud mental para menores en un centro residencial; (2) Program to Aid Citizen Enterprise (PACE): análisis de la equidad en la distribución de recursos a ONGs que sirven comunidades racializadas.
Contexto	El estudio se realizó en EE. UU., centrado en la región de Pittsburgh y en servicios de salud mental infantil. Aunque fue implementado por organizaciones sin ánimo de lucro, contó con el apoyo y financiación de agencias federales como HHS, HUD y la National Science Foundation.
Descripción general	El estudio aplica modelos causales con SCM y Precision Analytics para evaluar programas sociales. En Gemma Services se personalizan tratamientos de salud mental para menores; en PACE se analiza la distribución equitativa de fondos a ONGs. Ambos casos muestran cómo usar datos masivos para decisiones más equitativas y efectivas.
Tecnología aplicada	<ul style="list-style-type: none"> ▪ Structural Causal Modeling (SCM) y Precision Analytics (PA). ▪ Machine learning supervisado, análisis cuasi-experimental, matching de grupos. ▪ Dashboards interactivos, análisis geoespacial. ▪ Plataforma: Equitable Impact Platform (Equip) de BCT Partners.
Resultados documentados	<ul style="list-style-type: none"> ▪ Reducción de 30 días en la estancia media por niño en Gemma Services (ahorro estimado de \$10,000 por caso). ▪ Disminución del 40 % en el nivel de gravedad en el momento del alta, asociado a una menor tasa de rehospitalización. ▪ Identificación de combinaciones de tratamientos más efectivas para subgrupos específicos. ▪ En PACE, evidencia de desigualdad en la financiación de ONGs que atienden a comunidades racializadas.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Dependencia de la calidad y exhaustividad de los datos disponibles. ▪ Presencia de falsos positivos/negativos que requieren validación cualitativa. ▪ Riesgo de sesgos algorítmicos si no se controlan adecuadamente variables sociales. ▪ Beneficiarios aún no involucrados directamente en el uso de herramientas predictivas. ▪ Dificultades para interpretar y aplicar resultados sin capacitación técnica adecuada.
Referencia	York, P. (2024). The Future of Evaluation Analytics: Case Studies of Structural Causal Modeling in Action. En S. B. Nielsen, F. M. Rinaldi, & G. J. Petersson (Eds.), Artificial Intelligence and Evaluation (pp. 219-241). Routledge. https://doi.org/10.4324/9781003512493-11

11. Minería de texto y Machine Learning para evaluación temática (plurinacional)

Caso de uso	Minería de Texto y Machine Learning Supervisado para la evaluación temática de los indicadores relativos a los años 2010-20 del programa de préstamos Doing Business (DB) del Banco Mundial.
Contexto	La evaluación buscaba analizar la relevancia y efectividad de los indicadores DB. El reto consistía en identificar todos los proyectos influenciados por los indicadores DB entre 5,710 proyectos aprobados durante el período de evaluación.
Descripción general	El Independent Evaluation Group (IEG) desarrolló una metodología híbrida que combina NLP supervisado, búsquedas por palabras clave y revisión manual selectiva. Este enfoque iterativo permitió que los evaluadores aplicaran su juicio crítico a un subconjunto manejable de proyectos, mientras los algoritmos procesaban eficientemente grandes volúmenes de documentación.
Tecnología aplicada	Combinación de preprocesamiento lingüístico (eliminación de stopwords, lematización), vectorización mediante TF-IDF, ensemble de modelos de clasificación supervisada, y sistemas automatizados de búsqueda con puntuación de relevancia basada en frecuencia de términos para identificar eficientemente proyectos relevantes entre grandes volúmenes documentales.
Resultados documentados	<ul style="list-style-type: none"> ▪ 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 significativa del tiempo de evaluación al combinar la revisión humana con procesamiento automatizado. ▪ Los modelos de clasificación desarrollados para esta tarea fueron reutilizados en otras evaluaciones del IEG, maximizando el retorno de la inversión inicial.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Las limitaciones de recursos pueden restringir el tamaño y calidad de los datos de entrenamiento, afectando la precisión de los modelos de NLP supervisados. ▪ La diversidad de formas en que los proyectos pueden relacionarse con áreas regulatorias como "obtención de crédito" dificulta la captura completa mediante taxonomías de búsqueda. ▪ Es esencial mantener una comunicación estrecha entre científicos de datos y evaluadores para desarrollar un entendimiento compartido sobre la aplicación.
Referencia	Ziulu, V., Anuj, H., Hagh, A., Raimondo, E., & Vaessen, J. (2024). Extracting Meaning from Textual Data for Evaluation: Lessons from Recent Practice at the Independent Evaluation Group of the World Bank. En S. Bohni Nielsen, F. Mazzeo Rinaldi, & G. J. Petersson (Eds.), Artificial Intelligence and Evaluation: Emerging Technologies and Their Implications for Evaluation (1st ed., pp. 78-103). Routledge. https://doi.org/10.4324/9781003512493

12. Ingeniería de datos y Machine Learning para medir impacto en pobreza (plurinacional)

Caso de uso	Se utiliza un enfoque de ingeniería de datos y aprendizaje automático para procesar información digital y generar indicadores socioeconómicos georreferenciados que permiten evaluar intervenciones contra la pobreza.
Contexto	Colaboración interinstitucional entre organismos académicos, centros de investigación tecnológica y entidades multilaterales de cooperación al desarrollo.
Descripción general	Sistema de análisis computacional que combina datos de telefonía móvil (consumo de datos, actividad en redes sociales, transacciones financieras) e imágenes satelitales (usos de tierra, materiales de construcción, automóviles, iluminación nocturna, infraestructura urbana). Esta información es procesada mediante algoritmos de IA y aprendizaje automático para generar indicadores socioeconómicos georreferenciados.
Tecnología aplicada	Machine Learning aplicado al reconocimiento de patrones en datos heterogéneos, visión artificial para procesamiento de imágenes satelitales, y algoritmos de correlación para vincular indicadores digitales con métricas socioeconómicas tradicionales. Incluye técnicas de procesamiento de grandes volúmenes de datos (Big Data).
Resultados documentados	<ul style="list-style-type: none"> ▪ 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. ▪ Aumento en la granularidad espacial de los indicadores de pobreza. ▪ Acceso a medición en zonas con limitaciones logísticas o presupuestarias.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Sesgos muestrales por exclusión de población sin acceso a tecnología digital. ▪ Validación limitada respecto a mediciones socioeconómicas tradicionales. ▪ Eficacia diferencial según contexto: las imágenes satelitales son más efectivas en zonas rurales mientras que los datos telefónicos funcionan mejor en entornos urbanos. Distorsiones en la captura de asentamientos urbanos de alta densidad. ▪ Éticas: problemas potenciales de privacidad en el uso de datos personales; riesgo de amplificación de brechas digitales por sub-representación de poblaciones vulnerables.
Referencia	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, e3891. https://doi.org/10.15174/au.2023.3891

13. Automatización de informes de vulnerabilidad energética (Cataluña)

Caso de uso	Automatización mediante IA de los informes sociales de vulnerabilidad energética requeridos por la Ley 24/2015 para evitar cortes de suministros básicos a personas en riesgo de exclusión residencial.
Contexto	Estudio de caso sobre implementación de un sistema piloto en seis municipios catalanes: El Prat de Llobregat, Amposta, Granollers, Sant Boi de Llobregat, Palau Solità i Plegamans y Santa Coloma de Gramenet.
Descripción general	Sistema de automatización para generación de informes sociales de vulnerabilidad energética, requeridos mensualmente por las áreas básicas de servicios sociales (ABSS) para acreditar situaciones de vulnerabilidad económica ante posibles cortes de suministros por impago. El sistema integra múltiples fuentes de datos mediante una plataforma en la nube (alojada en Microsoft Azure-UE) que permite: cargar archivos de proveedores energéticos, obtener datos socioeconómicos de titulares, calcular coeficientes, generar informes automáticos de vulnerabilidad y comunicarlos a las empresas suministradoras a través de e-Notum.
Tecnología aplicada	Combinación de algoritmos de procesamiento de datos con tecnologías de Automatización Robótica de Procesos (RPA) y Gestión de Procesos de Negocio (BPM). Sistema desarrollado por la empresa privada Consultors BPM que integra servicios de interoperabilidad Via Oberta y el sistema de información de servicios sociales Hèstia.
Resultados documentados	<ul style="list-style-type: none"> ▪ Cobertura: 90% de éxito en la automatización de informes (con expectativa de alcanzar casi el 100% tras ajustes legales y organizativos). ▪ Reducción significativa de tiempo y cargas administrativas tanto para profesionales como ciudadanos. ▪ Interoperabilidad entre empresas, municipios y administraciones. ▪ Mejora en la gestión de datos y visión más precisa de situaciones de vulnerabilidad. ▪ Acceso a tecnología para municipios con recursos limitados mediante servicio en la nube sin necesidad de instalación local.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Necesidad de supervisión humana para validar resultados. ▪ Implicaciones de protección de datos al trabajar con información personal sensible. ▪ Requerimiento de integración con múltiples sistemas y compañías suministradoras (más de 50 en el territorio).
Referencia	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/

14. Decisiones automatizadas (ADM) en el sector público (Canadá)

Caso de uso	Análisis del impacto de la inteligencia artificial (IA) en los procesos de toma de decisiones del sector público y las transformaciones organizacionales resultantes desde una perspectiva de racionalidad limitada.
Contexto	Administración pública canadiense en la era digital y los sistemas de toma de decisiones automatizados.
Descripción general	Exploración de los sistemas de toma de decisiones automatizados (ADM) potenciados por IA en el sector público, analizando la relación entre elementos humanos y tecnológicos dentro de las estructuras organizativas gubernamentales. El estudio desarrolla un modelo de racionalidad limitada modernizado para examinar sistemas ADM basados en IA.
Tecnología aplicada	Sistemas de aprendizaje automático (Machine Learning), particularmente aprendizaje profundo (Deep Learning), redes neuronales, árboles de clasificación y máquinas de vectores de soporte, aplicados a sistemas automatizados de decisiones en el sector público.
Resultados documentados	<ul style="list-style-type: none"> ▪ Aumento significativo de la eficiencia organizacional al mejorar la evaluación de elementos factuales en las decisiones. ▪ Se prevé una transformación organizacional fundamental conforme la IA modifica las nociones tradicionales de "buena decisión" y las estructuras organizacionales.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Se identifican importantes desafíos éticos, incluyendo problemas de explicabilidad, sesgo algorítmico y disminución de la rendición de cuentas en sistemas de IA avanzados. ▪ Los sistemas de IA actuales son limitados para comprender plenamente los valores humanos complejos, lo que crea tensiones entre eficiencia técnica y juicio ético. ▪ Existe una brecha de conocimiento técnico en el sector público para implementar y supervisar sistemas de IA de manera efectiva.
Referencia	Alexander, W. (2022). Applying Artificial Intelligence to Public Sector Decision Making. Major Research Paper. University of Ottawa.

15. Experimentos naturales con datos históricos (plurinacional)

Caso de uso	Ensayo analítico con componentes teóricos y aplicados, basado en revisión conceptual y ejemplificación de casos.
Contexto	África y el contexto global de desarrollo; 2019-2020.
Descripción general	El artículo examina las implicaciones de cinco tecnologías disruptivas clave en el campo de la evaluación: 1) Big data, 2) Internet de las Cosas (IoT), 3) Drones y satélites, 4) Inteligencia artificial y minería de datos, y 5) Blockchain. Incluye un caso ilustrativo de una evaluación de impacto de un proyecto de construcción de carreteras en Ghana para demostrar aplicaciones potenciales.
Tecnología aplicada	Algoritmos de aprendizaje automático para: <ul style="list-style-type: none"> ▪ encontrar experimentos naturales en datos históricos para derivar conclusiones cuasi-experimentales sobre atribución/contribución de programas; ▪ automatizar procesos de evaluación; ▪ crear grupos de comparación emparejados mediante modelos predictivos; ▪ procesamiento de lenguaje natural para análisis cualitativo.
Resultados documentados	<ul style="list-style-type: none"> ▪ 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. ▪ Análisis más profundo de casos atípicos positivos. ▪ Evaluación de sostenibilidad a largo plazo. ▪ Se documenta un caso donde algoritmos de PNL con apoyo humano fue ocho veces más rápido que codificadores humanos sin pérdida de precisión.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Poca familiaridad con las tecnologías, resistencias de evaluadores veteranos, necesidad de reorganización y construcción de equipos. ▪ Inversiones significativas en equipos, software, personal y consultores son necesarias. ▪ Necesidad de reestructurar relaciones entre oficinas de evaluación, centros de datos y departamentos operativos. ▪ Riesgo de brecha digital entre organizaciones. ▪ Preocupaciones sobre privacidad, propiedad y uso ético de datos.
Referencia	Bamberger, M., & York, P. (2020). Transforming Evaluation in the 4th Industrial Revolution: Exciting Opportunities and New Challenges. <i>eVALUation Matters Second Quarter 2020</i> , 11-21. African Development Bank Group.

16. Machine Learning para evaluar el impacto del IMV (España)

Caso de uso	Aplicación de inteligencia artificial supervisada para evaluar el impacto del Ingreso Mínimo Vital (IMV) en el bienestar financiero de los hogares en España, distinguiendo entre efectos objetivos (ingreso, pobreza) y subjetivos (percepción de situación económica), para optimizar políticas públicas de inclusión social.
Contexto	Estudio sobre la implantación del IMV en España, desarrollado en el marco académico de la London School of Economics.
Descripción general	Se emplea inteligencia artificial supervisada, concretamente el Método de Control Sintético (SCM) y su versión mejorada, el Ridge Augmented SCM (RASCAM). Estas técnicas, consideradas dentro del campo del machine learning supervisado, permiten generar un contrafactual automatizado (España sin IMV) para estimar de forma causal el efecto real del IMV, evitando suposiciones conductuales y corrigiendo sesgos de selección.
Tecnología aplicada	Algoritmos de IA entrenados con datos panel nacionales (2010–2022) de Eurostat. Indicadores analizados: tasa de pobreza, brecha de pobreza, ingreso medio y percepción subjetiva. Comparación con un grupo de control sintético formado por 11 países de la UE.
Resultados documentados	<ul style="list-style-type: none"> ▪ Sin efectos significativos en indicadores objetivos durante el primer año y medio. ▪ A los 2,5 años, mejora perceptible en el bienestar subjetivo (aumento de 10–14 puntos en percepción positiva). ▪ Indicadores como brecha de pobreza y renta media mejoran levemente, pero sin significancia estadística. ▪ La IA permite captar impactos diferenciales en dimensiones objetivas y percibidas del bienestar.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Limitado impacto estadístico en los indicadores objetivos evaluados. ▪ Retrasos en la implementación y baja cobertura del IMV reducen su efecto medido. ▪ Las percepciones están sujetas a sesgos individuales y contexto socioeconómico externo (COVID, inflación). ▪ Dependencia de la calidad y disponibilidad de datos comparables entre países.
Referencia	Bilbao-Goyoaga, E. (2023) Perceptions Matter: Quasi-Experimental Evidence on the Effects of Spain's New Minimum Income on Households' Financial Wellbeing, <i>Social Policy Working Paper 02-23</i> , London: 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

17. Modelo de Lenguaje Grande (LLM) en evaluación (estudio analítico)

Caso de uso	Ensayo analítico y exploratorio que combina reflexión conceptual con ejemplos prácticos de aplicación. El artículo explora las posibilidades y desafíos del uso de ChatGPT como herramienta de apoyo en diferentes etapas del proceso de evaluación, desde la generación de términos de referencia hasta la diseminación de informes finales.
Contexto	Sin especificación geográfica concreta, aunque con perspectiva global sobre la disciplina evaluativa. 2022-2023: periodo inicial de disponibilidad pública de ChatGPT (publicado en 2023, con experiencias basadas en GPT-3).
Descripción general	La autora examina el uso de "hacking by the prompt" (uso creativo y hábil de indicaciones conversacionales) para interactuar con ChatGPT en diversos aspectos del trabajo evaluativo. A través de ejemplos y reflexiones críticas, se muestran tanto las posibilidades como las limitaciones de integrar IA en la práctica evaluativa. Se incluyen ejemplos concretos de prompts (indicaciones) para cada etapa del proceso evaluativo.
Tecnología aplicada	Modelo de lenguaje grande (LLM) ChatGPT-3 y referencias a ChatGPT-4, una IA generativa pre-entrenada con capacidades de procesamiento y generación de lenguaje natural, accesible a través de interfaz conversacional.
Resultados documentados	El estudio documenta varios usos potenciales en evaluación: <ul style="list-style-type: none"> ▪ Generación de documentos técnicos (términos de referencia, notas conceptuales, marcos lógicos). ▪ Diseño de instrumentos (cuestionarios, guías de entrevista). ▪ Análisis de texto (resúmenes, extracción de palabras clave, análisis de sentimiento). ▪ Apoyo a la comunicación (adaptación de contenido para audiencias específicas, traducción).
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ ChatGPT opera con probabilidades, no con conocimiento factual verificado. ▪ Produce "alucinaciones" (información fabricada que parece plausible pero es falsa). ▪ La falta de transparencia sobre los datos de entrenamiento genera preocupaciones sobre sesgos y visiones del mundo predominantes. Preocupaciones sobre privacidad de datos y ética. ▪ Riesgo de reemplazar el pensamiento crítico y el juicio humano por eficiencia automatizada. ▪ Falta de control sobre la evolución y regulación de los sistemas de IA.
Referencia	Ferretti, S. (2023). Hacking by the prompt: Innovative ways to utilize ChatGPT for evaluators. <i>New Directions for Evaluation</i> , 2023, 73-84. https://doi.org/10.1002/ev.20557

18. Machine Learning para una evaluación basada en la teoría del cambio (plurinacional)

Caso de uso	Aplicación de inteligencias artificiales para automatizar y mejorar el análisis de contenido y la síntesis cualitativa en evaluaciones de programas complejos basados en teorías de cambio.
Contexto	Evaluación temática del Independent Evaluation Group (IEG) del Banco Mundial sobre programas de reducción de desnutrición infantil crónica y retraso en el crecimiento: 392 informes de proyectos de 64 países.
Descripción general	Implementación de tres enfoques complementarios de IA: (1) aprendizaje automático supervisado (SML) para clasificación deductiva de textos según un marco conceptual predefinido; (2) aprendizaje automático no supervisado (UML) para identificar temas emergentes inductivos de factores de éxito y fracaso; (3) grafos de conocimiento para estructurar los resultados según una teoría del cambio y facilitar el análisis basado en dicha teoría.
Tecnología aplicada	<ul style="list-style-type: none"> ▪ Modelado de espacios vectoriales para representación numérica de textos. ▪ Aprendizaje automático supervisado (clasificadores de texto) y no supervisado (modelado de temas). ▪ Visualización t-SNE (t-distributed stochastic neighbor embedding). ▪ Grafos de conocimiento con razonamiento basado en reglas (Vadalog).
Resultados documentados	<ul style="list-style-type: none"> ▪ SML: precisión modesta en la predicción de etiquetas exactas; alta precisión en la predicción de categorías de nivel superior; capacidad para identificar características distintivas de programas por país mediante visualización t-SNE. ▪ UML: excelente rendimiento en la identificación de 10 temas coherentes y relevantes como factores clave que influyen en el desempeño del proyecto; validación estadística posterior. ▪ Grafos de conocimiento: capacidad para estructurar outputs de aprendizaje automático según la teoría de cambio; identificación de relaciones simples entre intervenciones y logros de indicadores de resultados.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Inversión significativa de tiempo inicial para preparar datos y establecer marcos analíticos y esquema de grafos. ▪ Dependencia de expertos para extraer manualmente secciones de texto relevantes para etiquetado. ▪ Evidencia incompleta vinculando elementos del marco conceptual (algunos nodos quedan "huérfanos").
Referencia	Franzen, S. et al. (2022). Advanced Content Analysis: Can Artificial Intelligence Accelerate Theory-Driven Complex Program Evaluation? <i>IEG Methods and Evaluation Capacity Development Working Paper Series</i> . Independent Evaluation Group. Washington, DC: World Bank.

19. Personas evaluativas: NLP y modelos de aprendizaje automático (Camboya)

Caso de uso	Introducción y aplicación de "personas evaluativas" como herramienta metodológica para analizar datos cualitativos complejos en evaluaciones de programas y generar recomendaciones accionables.
Contexto	Programa de saneamiento en Camboya implementado por iDE y financiado por Australian Aid a través del programa Water for Women.
Descripción general	El artículo explora el uso de “personas” evaluativas, es decir, perfiles ficticios basados en datos reales, como herramienta para mejorar el análisis y la utilidad de evaluaciones cualitativas en programas de desarrollo. En el caso estudiado, un equipo de investigación aplicó este enfoque durante la evaluación del impacto que un programa de transversalización de género había generado en los miembros del personal (staff) de un programa de saneamiento en Camboya. A partir de 199 micro-narrativas recogidas de 176 trabajadores, se crearon 14 personas.
Tecnología aplicada	Procesamiento de Lenguaje Natural (NLP) mediante dos algoritmos: Latent Dirichlet Allocation (LDA) y Structural Topic Modeling (STM), utilizados para validar los grupos de personas identificados mediante análisis temático cualitativo manual.
Resultados documentados	<ul style="list-style-type: none"> Identificación de 14 personas agrupadas en tres categorías teóricas: observadores críticos (34%), auto-reflexivos críticos (18%) y actores críticos (48%). Las personas permitieron visualizar y comunicar efectivamente datos cualitativos complejos. Facilitaron discusiones con líderes del programa sobre estrategias de incorporación de género y proporcionaron cierto anonimato.
Riesgos y limitaciones	<ul style="list-style-type: none"> En la recolección de micro-narrativas retrospectivas, los investigadores no pueden contrastar la validez de las historias, lo que puede incluir cierto grado de embellecimiento; cierta distancia (física y temporal) del programa. El enfoque de muestreo censal creó problemas de anonimato que las personas ayudaron a mitigar, pero que siguen siendo una consideración ética importante.
Referencia	MacArthur, J., Moung, V., Carrard, N., & Willetts, J. (2024). Personas for program evaluation: Insights from a gender-focused evaluation in Cambodia. <i>Evaluation</i> , 31(1), 70-91. https://doi.org/10.1177/13563890241284425

20. Big data para detectar tendencias en desempeño (Israel)

Caso de uso	Desarrollo e implementación de una metodología automatizada basada en minería de textos, aprendizaje automático y visualización de datos para detectar tendencias en grandes volúmenes de informes de desempeño y evaluación no estructurados.
Contexto	El sistema se aplicó para analizar 1700 informes de auditoría y evaluación de desempeño producidos por la Intervención del Estado de Israel (State Comptroller) durante un período de 30 años (1987-2017).
Descripción general	<p>El método utiliza Wikipedia y otras fuentes externas como bases de conocimiento para identificar términos relevantes en los documentos. Mediante algoritmos de aprendizaje automático no supervisado, el sistema crea una representación visual en forma de gráfico donde los informes semánticamente relacionados aparecen agrupados en clusters. Los diferentes colores en la visualización representan los períodos de servicio de distintos interventores del Estado.</p> <p>Esta metodología permitió a los investigadores identificar patrones y tendencias que habrían sido prácticamente imposibles de detectar mediante métodos tradicionales de análisis de contenido.</p>
Tecnología aplicada	Análisis estadístico de lenguaje natural (SNLP), aprendizaje automático no supervisado para clasificación de textos, análisis semántico explícito (ESA), reducción de dimensionalidad no lineal, y algoritmos de visualización de grafos (Force Atlas y Fruchterman-Reingold).
Resultados documentados	El sistema identificó eficazmente clusters temáticos, trayectorias temporales y relaciones entre documentos sin necesidad de búsquedas por palabras clave predefinidas. Se revelaron patrones como: (1) la transición de auditorías tradicionales hacia auditorías de desempeño, (2) cambios en los enfoques de diferentes interventores del Estado, (3) continuidad/discontinuidad en temas evaluados, y (4) conexiones entre clusters temáticos. El enfoque permitió reducir la revisión de 1700 informes a aproximadamente 50 informes representativos para obtener una visión global del dominio.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ El sistema requiere programación personalizada para cada campo. Necesita ajustes específicos que implican conocimientos técnicos, creando una barrera de entrada para usuarios sin experiencia en desarrollo. ▪ Aunque evita el sesgo de búsqueda por palabras clave, introduce otros posibles sesgos a través de decisiones como usar Wikipedia como base de conocimiento o la selección de algoritmos específicos de visualización que afectan la representación de datos.
Referencia	Raveh, E., Ofek, Y., Bekkerman, R., & Cohen, H. (2020). Applying Big Data visualization to detect trends in 30 years of performance reports. <i>Evaluation</i> , 26(4), 516-540. https://doi.org/10.1177/1356389020905322

21. Case Based Modelling (CBM) y simulación de escenarios (Reino Unido)

Caso de uso	Desarrollo y aplicación de una metodología de modelado basado en casos y simulación de escenarios (CBSS) para evaluar intervenciones políticas en sistemas complejos del mundo real.
Contexto	Analiza cuatro evaluaciones publicadas de programas del Departamento de Negocios, Energía y Estrategia Industrial (BEIS) y el Departamento de Educación (DfE) del gobierno británico entre 2008 y 2018, y en concreto: Evaluación piloto del Heat Network Investments Project (HNIP); Evaluación del Renewable Heat Incentive (RHI) – BEIS; Evaluación del programa piloto Frontline Fast-Track Social Work; Evaluación Nacional de Sure Start – DfE.
Descripción general	CBSS (Case-Based Scenario Simulation) es una herramienta de investigación social para evaluación ex-post, basada en un enfoque de modelado computacional de sistemas basado en casos y orientado a datos. Implementada a través de la plataforma web COMPLEX-IT, permite a los evaluadores identificar patrones de casos mediante análisis de conglomerados (clustering), corroborar resultados con mapas auto-organizados (SOM) y simular escenarios alternativos para explorar contrafactuales y proyecciones.
Tecnología aplicada	Aplicación de métodos de aprendizaje automático no supervisado, específicamente: (1) análisis de conglomerados k-means para identificar grupos de casos, (2) redes neuronales artificiales a través de mapas auto-organizados de Kohonen (SOM) para visualización topográfica de casos, y (3) simulaciones Monte Carlo para análisis de sensibilidad.
Resultados documentados	<ul style="list-style-type: none"> ▪ Capacidad demostrada para modelar sistemas complejos sin reducir la explicación a factores individuales. ▪ Identificación de subgrupos con diferentes respuestas a intervenciones. ▪ Exploración de contrafactuales y escenarios alternativos. ▪ Evaluación de la robustez de soluciones de agrupamiento. ▪ Complementariedad con enfoques de evaluación realista.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Limitaciones para abordar las incertidumbres en sistemas complejos. ▪ No identifica relaciones causales definitivas, sino que proporciona un entorno para explorarlas. ▪ Requiere especificación adecuada de casos y perfiles de casos. ▪ Sensible a la calidad y completitud de los datos. ▪ Dependiente de la interacción constante del evaluador con el entorno de aprendizaje.
Referencia	Schimpf, C., Barbrook-Johnson, P., & Castellani, B. (2021). Cased-based modelling and scenario simulation for ex-post evaluation. <i>Evaluation</i> , 27(1), 116-137. https://doi.org/10.1177/1356389020978490

22.Chatbots para la formación de evaluadores (EE. UU. y Canadá)

Caso de uso	Exploración de las implicaciones y aplicaciones potenciales de la Inteligencia Artificial en la formación de evaluadores, incluyendo prototipos de chatbots para mejorar la experiencia formativa en el campo de la evaluación.
Contexto	Ámbito académico y educativo en programas de formación en evaluación, principalmente en Estados Unidos y Canadá, durante 2022-2023, con perspectivas de aplicación global.
Descripción general	Desarrollo de tres prototipos de chatbots basados en IA generativa para la formación en evaluación: (1) un orientador para responder preguntas sobre programas formativos en evaluación, (2) un apoyo docente para aclarar teorías específicas de evaluación, y (3) un mentor para responder preguntas fundamentales sobre la práctica evaluativa.
Tecnología aplicada	Inteligencia Artificial generativa basada en Modelos de Lenguaje Grande (LLM), específicamente utilizando el modelo GPT 3.5 turbo de OpenAI con Procesamiento de Lenguaje Natural (NLP), implementado mediante Generación Aumentada por Recuperación (RAG) para proporcionar respuestas basadas en corpus documentales específicos sobre evaluación.
Resultados documentados	Los prototipos demostraron capacidad para: <ul style="list-style-type: none"> ▪ proporcionar información precisa sobre programas formativos sobre evaluación en EE.UU.; ▪ sintetizar respuestas a preguntas teóricas combinando múltiples documentos; ▪ responder preguntas sobre práctica evaluativa con un enfoque adaptado a evaluadores emergentes.
Riesgos y limitaciones	<ul style="list-style-type: none"> ▪ Técnicas: variabilidad en la calidad de traducción automática multilingüe, particularmente para terminología especializada de evaluación; limitaciones en la ventana de contexto de los modelos utilizados; riesgo de alucinaciones o respuestas incorrectas. ▪ Metodológicas: evaluación preliminar y no sistemática del rendimiento de los chatbots; ausencia de métricas formales para evaluar la precisión y utilidad de las respuestas generadas; limitada prueba con usuarios reales. ▪ Éticas: posibilidad de perpetuar sesgos y exclusiones históricas en investigación y evaluación a través de los modelos de IA; preocupaciones sobre la equidad de acceso a estas tecnologías, especialmente en el Sur Global.
Referencia	Tilton, Z., LaVelle, J. M., Ford, T., & Montenegro, M. (2023). Artificial intelligence and the future of evaluation education: Possibilities and prototypes. <i>New Directions for Evaluation</i> , 2023(178-179), 97-109. https://doi.org/10.1002/ev.20564



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