Enpath

INFORME · INVESTIGACIÓN

Supervisión humana en las decisiones públicas: lo que nunca debe delegarse

Febrero de 2024

Organización independiente dedicada a promover un uso responsable de la inteligencia artificial en las decisiones públicas y organizacionales. Cada publicación incluye su metodología, sus fuentes y sus limitaciones.

Enpath · **Investigación** · Investigación con marco propio · Pilar: Supervisión humana **Publicado:** febrero de 2024 · **Corte de información:** 29 de febrero de 2024 · **Idioma:** español

> Los mayores desastres documentados de la automatización pública —Robodebt en Australia, los subsidios en Países Bajos, las calificaciones en el Reino Unido, los procesamientos del caso Horizon— comparten una causa: en algún punto, las personas dejaron de decidir y empezaron a firmar lo que el sistema producía. Este documento examina esa evidencia y propone dos herramientas: una lista de decisiones que nunca deben delegarse y una matriz para ubicar todas las demás.

Actualización.

Actualizado el: 18/07/2026. Esta sección resume acontecimientos posteriores a la fecha original de publicación y no forma parte del análisis histórico original.

- La Corte Constitucional de Colombia profirió en agosto de 2024 la sentencia T-323 de 2024, a propósito del uso de ChatGPT por un juez de Cartagena que este informe analiza. La Corte fijó criterios para el uso de inteligencia artificial en la Rama Judicial: la IA no puede sustituir el razonamiento del juez y su uso exige transparencia, responsabilidad y protección de datos.
- El Reglamento de IA de la Unión Europea fue adoptado y entró en vigor el 1 de agosto de 2024; su artículo 14 (supervisión humana), analizado aquí sobre el texto acordado, quedó en la versión final con obligaciones para sistemas de alto riesgo aplicables desde 2026.
- El Reino Unido expidió en mayo de 2024 la ley que anuló en bloque las condenas del caso Horizon (Post Office (Horizon System) Offences Act 2024).
- En Estados Unidos, el memorando OMB M-24-10 (marzo de 2024) hizo obligatorias salvaguardas —incluida supervisión humana— para usos de IA que afectan derechos en agencias federales.

Resumen ejecutivo.

Delegar una decisión no es lo mismo que apoyarla. Un sistema de inteligencia artificial puede ordenar expedientes, detectar patrones y sugerir priorizaciones; lo que no puede hacer —ni técnica ni jurídicamente— es responder por la decisión. Cuando la cadena de decisión pública pierde a la persona que examina el caso concreto, el error deja de ser una excepción corregible y se convierte en política institucional ejecutada a escala.

La evidencia es abundante y reciente. En Australia, un esquema automatizado de recuperación de deudas emitió más de 470.000 cobros erróneos; la Comisión Real que lo investigó lo calificó en 2023 de "mecanismo burdo y cruel, ni justo ni legal" (Royal Commission, 2023). En Países Bajos, un sistema de detección de fraude en subsidios señaló injustamente a cerca de 26.000 familias y contribuyó a la caída

del gobierno en 2021. En el Reino Unido, el 39,1 % de las calificaciones de acceso a la universidad fueron rebajadas por un algoritmo en 2020 (Ofqual, 2020); y el caso Horizon —más de 900 procesamientos penales sustentados en un sistema informático defectuoso— dominaba el debate público británico al cierre de este informe.

Este documento extrae de esa evidencia una regla y dos herramientas. La regla: **la supervisión humana no se declara; se diseña, se dota de recursos y se mide.** Las herramientas: una lista de seis decisiones que no deben delegarse en ningún caso, y la **Matriz Enpath de delegación**, que ubica cualquier otra decisión según su impacto sobre derechos y la reversibilidad del error. El marco se apoya en los instrumentos vigentes —el artículo 22 del RGPD europeo, el artículo de supervisión humana del Reglamento de IA acordado en la UE, la Directiva canadiense de decisiones automatizadas y el principio de "control humano de las decisiones" del Marco Ético colombiano— y en la investigación empírica sobre sesgo de automatización.

Hallazgos principales.

1. **El patrón de las fallas es institucional, no técnico.** En Robodebt, Toeslagenaffaire, Ofqual y Horizon, los sistemas hicieron aquello para lo que fueron configurados. Lo que falló fue la arquitectura de decisión: nadie con autoridad revisaba el caso individual, y quienes alertaron fueron ignorados.
2. **La supervisión nominal es la modalidad más frecuente de falla.** La investigación sobre sesgo de automatización muestra que las personas tienden a aceptar las recomendaciones de un sistema aun cuando contradicen otra evidencia disponible (Skitka, Mosier y Burdick, 1999); en contextos de decisión asistida, los revisores siguen las sugerencias de forma desigual y sesgada según el caso (Green y Chen, 2019). Poner a una persona "en el circuito" sin tiempo, información ni autoridad produce firmas, no supervisión.
3. **La regulación converge hacia la supervisión obligatoria.** El RGPD reconoce desde 2018 el derecho a no ser objeto de decisiones puramente automatizadas con efectos significativos (art. 22); el Reglamento de IA europeo acordado exige supervisión humana efectiva para sistemas de alto riesgo (los Estados miembros respaldaron el texto final el 2 de febrero de 2024); Canadá gradúa desde 2019 la intervención humana requerida según el nivel de impacto de cada sistema.
4. **Colombia ya tiene el principio; le falta el procedimiento.** El Marco Ético de 2021 consagra el "control humano de las decisiones" y la Constitución exige debido proceso y motivación de los actos administrativos. Pero ninguna norma colombiana especifica, a la fecha, cómo debe organizarse esa supervisión: quién revisa, con qué información y con qué registro. El uso de ChatGPT en una sentencia de Cartagena en 2023 —bajo revisión de la Corte Constitucional al cierre de este informe— muestra que la pregunta ya llegó a los despachos.
5. **Existen seis decisiones que ningún sistema debe tomar**, con independencia de su precisión: la decisión final sobre derechos fundamentales, las sanciones, el uso de la fuerza, las decisiones judiciales de fondo, los casos atípicos que el sistema no fue diseñado para tratar, y la asignación de responsabilidad. La sección central de este documento las desarrolla.

Por qué importa.

La decisión que este documento apoya es de diseño institucional: **al adoptar un sistema que participa en decisiones sobre personas, ¿qué nivel de automatización es admisible y cómo debe organizarse la supervisión humana para que sea real?**

La pregunta es urgente por una razón práctica: la supervisión humana es la salvaguarda a la que todos los marcos apelan —de la UNESCO al Marco Ético colombiano— y, a la vez, la que con más frecuencia existe solo en el papel. Cuando la supervisión es nominal, la entidad obtiene lo peor de ambos mundos: la lentitud del procedimiento humano y los errores a escala del sistema, sin la protección de ninguno de los dos.

Antecedentes.

El sesgo de automatización: lo que muestra la investigación.

La confianza excesiva en sistemas automatizados es uno de los fenómenos mejor documentados de la interacción humano-computador. Los experimentos clásicos de Skitka y colegas mostraron que operadores con apoyo automatizado cometen dos tipos de error: omisión (no detectar fallas que el sistema no señala) y comisión (seguir recomendaciones incorrectas contradichas por otra información disponible) (Skitka, Mosier y Burdick, 1999). En contextos de política pública, Green y Chen encontraron que evaluadores humanos que reciben puntajes de riesgo los siguen de manera inconsistente y con sesgos propios, sin lograr los niveles de precisión del sistema ni los de un juicio humano informado (Green y Chen, 2019). La investigación reciente en interacción humano-IA explora "funciones de forzamiento cognitivo" que obligan al revisor a examinar el caso antes de ver la recomendación, con resultados prometedores pero costos de tiempo (Buçinca, Malaya y Gajos, 2021).

Tres conclusiones prácticas se desprenden: la presencia de una persona no garantiza juicio; el diseño del flujo de trabajo determina si la persona puede ejercerlo; y la tasa de aceptación de recomendaciones es el indicador más simple para detectar supervisión nominal.

Cuatro fallas públicas, un mismo patrón.

Caso	Jurisdicción	Magnitud	Desenlace
Robodebt (2016–2020)	Australia	Más de 470.000 deudas emitidas erróneamente contra beneficiarios de asistencia social	Acuerdo judicial de AUD 1.800 millones (2021); la Comisión Real concluyó en 2023 que el esquema fue ilegal (Royal Commission, 2023)
Toeslagenaffaire (2013–2019)	Países Bajos	Cerca de 26.000 familias acusadas injustamente de fraude en subsidios de cuidado infantil	Renuncia del gobierno en pleno (enero de 2021); multa de EUR 2,75 millones a la administración tributaria por tratamiento discriminatorio de la doble nacionalidad (Autoriteit Persoonsgegevens, 2021)
Calificaciones Ofqual (2020)	Reino Unido		Algoritmo retirado en una semana; se usaron las evaluaciones docentes (Ofqual, 2020)

Caso	Jurisdicción	Magnitud	Desenlace
		39,1 % de las calificaciones A-level rebajadas frente a la evaluación docente	
Horizon / Post Office (1999–2015)	Reino Unido	Más de 900 procesamientos penales sustentados en datos de un sistema contable defectuoso	93 condenas anuladas a enero de 2024; el Gobierno anunció una ley de exoneración general (Gobierno del Reino Unido, 2024)
SyRI (2014–2020)	Países Bajos	Perfilamiento de riesgo de fraude concentrado en barrios de bajos ingresos	El Tribunal de La Haya lo declaró contrario al art. 8 del Convenio Europeo de Derechos Humanos (febrero de 2020) (Rechtbank Den Haag, 2020)

Tabla 1. Cinco fallas documentadas de decisión automatizada o asistida en el sector público. Fuente: elaboración de Enpath con base en los documentos citados.

El patrón común no es la mala fe del algoritmo sino la evaporación de la responsabilidad: en cada caso, la organización trató la salida del sistema como si fuera una decisión ya tomada. En Horizon, además, operó una presunción legal de fiabilidad de la evidencia informática que invirtió la carga de la prueba contra las personas. La lección es exactamente la inversa de la que suele extraerse: el problema no fue usar tecnología, sino dejar de decidir.

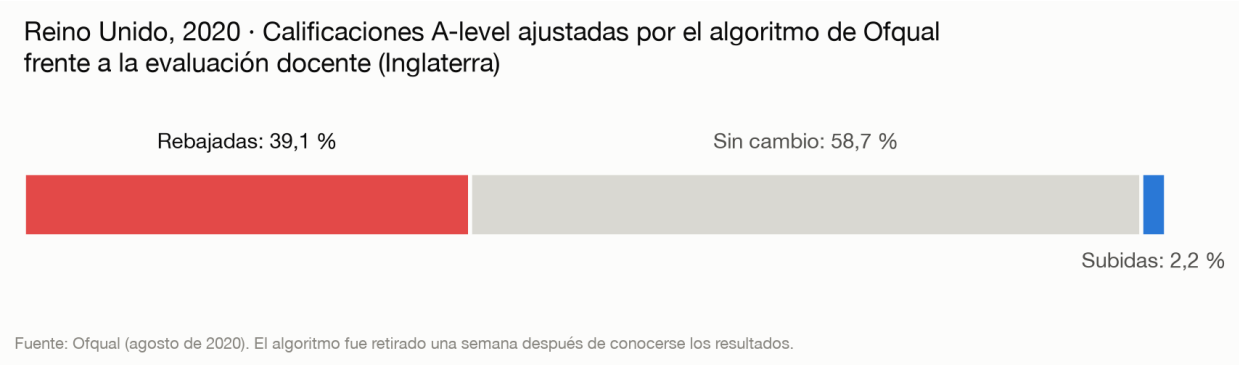

Figura 1. Reino Unido, 2020: proporción de calificaciones A-level ajustadas por el algoritmo de Ofqual frente a la evaluación docente en Inglaterra. Fuente: Ofqual (2020).

Lo que nunca debe delegarse.

Seis categorías de decisión deben permanecer en manos humanas con independencia de la calidad del sistema. No son una preferencia de diseño: se derivan del debido proceso, de la motivación exigible a los actos públicos y de la imposibilidad jurídica de trasladar la responsabilidad a una máquina.

1. **La decisión final sobre derechos fundamentales.** Acceso a salud, libertad personal, ciudadanía, asilo, custodia. El sistema puede ordenar información; la decisión y su motivación son humanas.
2. **Las sanciones.** Multas, suspensiones, inhabilidades, exclusión de beneficios. Toda sanción exige valoración individual de circunstancias, que es precisamente lo que la automatización a

escala elimina. Robodebt es el caso límite: cobros coactivos calculados por promediación de ingresos, sin revisión individual.

3. **El uso de la fuerza.** Ninguna aplicación policial, militar o de control migratorio debe ejecutar fuerza física o coerción sobre la base de una salida algorítmica sin decisión humana individual.
4. **Las decisiones judiciales de fondo.** Los sistemas pueden apoyar la gestión de despachos — como PretorIA en la selección de tutelas— pero la valoración probatoria, la interpretación jurídica y la decisión pertenecen al juez. El uso de un modelo de lenguaje para redactar apartes de una sentencia en Cartagena en 2023 planteó exactamente esta frontera, hoy bajo examen de la Corte Constitucional.
5. **Los casos atípicos.** Todo sistema tiene un dominio de diseño. Los casos que se apartan de él — el ciudadano sin historial, la situación no prevista, el dato faltante— deben derivarse a una persona, no forzarse dentro del modelo. La proporción de casos derivados es un indicador de honestidad del sistema: si es cero, el sistema está tratando casos que no entiende.
6. **La asignación de responsabilidad.** Ningún diseño puede hacer que "lo decidió el sistema" sea una respuesta válida. La responsabilidad se asigna antes de operar: un cargo, un nombre, un mecanismo de revisión.

La matriz de delegación.

Fuera de las seis categorías anteriores, el nivel de automatización admisible depende de dos variables que cualquier entidad puede evaluar: el **impacto sobre derechos** de la decisión y la **reversibilidad del error**.

Figura 2. Matriz Enpath de delegación. Fuente: Enpath.

La lógica de la matriz coincide con la de la Directiva sobre Decisiones Automatizadas de Canadá, que desde 2019 clasifica los sistemas federales en cuatro niveles de impacto y exige intervención humana creciente —hasta la aprobación humana específica en los niveles altos—, junto con evaluaciones de impacto algorítmico publicadas. La diferencia de la matriz de Enpath es su segunda dimensión: la reversibilidad. Un error masivo pero corregible en días (una notificación mal enviada) admite más automatización que un error individual de consecuencias difíciles de deshacer (una prestación negada durante meses a quien dependía de ella).

Figura 3. Tres modos de intervención humana. Fuente: Enpath, con base en las categorías del debate regulatorio europeo y canadiense.

Condiciones de la supervisión efectiva.

Declarar que "una persona revisa" no crea supervisión. Cinco condiciones, todas verificables, la crean:

Condición	Qué exige	Cómo se verifica
Competencia	El revisor entiende el dominio y las limitaciones del sistema	Perfil del cargo; formación documentada
Información	El revisor ve el caso completo, no solo la recomendación	Diseño de la interfaz; acceso al expediente
Tiempo	La carga de trabajo permite examinar cada caso	Casos por revisor por día, comparado con un estándar
Autoridad	Apartarse de la recomendación no tiene costo personal	Procedimiento escrito; ausencia de metas de "aceptación"
Registro	Cada intervención queda documentada	Bitácora de decisiones: acepta, modifica, rechaza, deriva

Tabla 2. Condiciones de la supervisión humana efectiva. Fuente: Enpath.

Dos métricas resumen el estado de la supervisión en cualquier sistema en operación:

- **Tasa de aceptación:** proporción de recomendaciones del sistema adoptadas sin modificación. Sostenida cerca del 100 %, indica supervisión nominal; el sistema decide y la persona firma.
- **Tasa de derivación:** proporción de casos que el flujo remite a decisión humana plena por atipicidad. Sostenida en cero, indica que el sistema procesa casos fuera de su dominio de diseño.

Ninguna de las dos métricas requiere tecnología adicional; requieren voluntad de medir.

Colombia.

El ordenamiento colombiano ya contiene las piezas jurídicas: el debido proceso es derecho fundamental de aplicación directa (Constitución, art. 29), los actos administrativos deben motivarse (Ley 1437 de 2011) y el Marco Ético para la IA (2021) adoptó expresamente el principio de "control humano de las decisiones" junto con la prevalencia de los derechos (Presidencia de la República, 2021). Lo que falta es el puente operativo: ninguna norma especifica cómo debe organizarse la supervisión —condiciones, métricas, registro— en las entidades que ya usan sistemas automatizados.

Dos casos ilustran los extremos del espectro colombiano:

- **PretorIA** (Corte Constitucional) es un ejemplo de delegación bien acotada: el sistema detecta categorías predefinidas en expedientes de tutela para apoyar la selección, los despachos deciden, y el proyecto se documentó públicamente con sus criterios (BID, 2021). La función del sistema es informativa; la decisión permanece en el juez.
- **La sentencia de Cartagena (enero de 2023)**, en la que un juez incorporó respuestas de ChatGPT en la motivación de una tutela, abrió el debate opuesto: el uso de un sistema generativo —no diseñado para el derecho colombiano, sin trazabilidad de fuentes— en la zona que este documento clasifica como indelegable. La Corte Constitucional seleccionó el caso para revisión, pendiente de decisión al cierre de este informe. Cualquiera que sea el resultado, el caso demostró que la frontera de delegación ya no es teórica en Colombia.

Para las entidades del Ejecutivo, la implicación es directa: los sistemas en uso —asociación de casos en la Fiscalía, asistentes de trámites, priorización en programas sociales— operan hoy sin métricas públicas de aceptación ni derivación. Adoptar las cinco condiciones y las dos métricas de este documento no requiere ley nueva; requiere un acto administrativo y disciplina de gestión.

América Latina.

El principio de supervisión humana aparece en todos los marcos de la región, con niveles distintos de exigibilidad. El Marco Ético colombiano (2021) y la Política Nacional chilena (2021) lo consagran como principio; la Ley 31814 de Perú (2023) declara el uso "responsable" de la IA sin desarrollar aún el mecanismo; el proyecto de ley brasileño 2338/2023 —en trámite en el Senado al cierre de este informe — incluye derechos de revisión humana de decisiones automatizadas con efectos significativos, siguiendo el modelo europeo. En paralelo, la experiencia acumulada por los programas de la banca

regional (fAIr LAC del BID) documenta que los pilotos con mejor desempeño institucional son los que definieron desde el diseño quién decide y qué hace el sistema (BID, 2021).

La región tiene, además, una razón estructural para tomarse en serio la supervisión: las capacidades de auditoría y control son más débiles que en la OCDE, de modo que el error automatizado tarda más en detectarse y corregirse. Donde la corrección es lenta, la matriz de delegación exige más decisión humana, no menos.

Comparación internacional.

Instrumento	Mecanismo de supervisión	Estado a feb. 2024
RGPD, art. 22 (UE, 2016)	Derecho a no ser objeto de decisiones puramente automatizadas con efectos jurídicos o significativos; excepciones con salvaguardas	Vigente y aplicado por autoridades de datos
Reglamento de IA, texto acordado (UE)	Supervisión humana efectiva obligatoria para sistemas de alto riesgo (art. 14): medidas proporcionales al riesgo, capacidad real de intervenir o detener el sistema	Texto final respaldado por los Estados miembros (2 de febrero de 2024); adopción formal pendiente
Directiva sobre Decisiones Automatizadas (Canadá, 2019)	Cuatro niveles de impacto; intervención humana y salvaguardas crecientes; evaluación de impacto algorítmico pública	Vigente; actualizada en 2023
Orden Ejecutiva 14110 y proyecto de memorando OMB (EE. UU., 2023)	Salvaguardas mínimas para IA que afecta derechos, incluida consideración explícita de supervisión y vías de reclamo	Orden vigente; memorando en proyecto
Marco Ético para la IA (Colombia, 2021)	"Control humano de las decisiones" como principio para proyectos públicos	Vigente como guía; sin mecanismo de verificación

Tabla 3. Mecanismos de supervisión humana en cinco instrumentos. Fuente: elaboración de Enpath con base en los documentos oficiales citados.

La comparación deja una conclusión incómoda para la región: los marcos latinoamericanos enuncian el principio con tanta claridad como los europeos, pero ninguno —a la fecha— lo convierte en condiciones verificables. La distancia entre el artículo 14 del reglamento europeo y el principio colombiano no está en la ambición sino en la maquinaria: umbrales, evaluaciones previas, autoridad de supervisión y sanción.

Oportunidades.

- **La supervisión efectiva es más barata que el litigio.** El costo del acuerdo de Robodebt (AUD 1.800 millones) financiaría décadas de revisión humana competente. Para una entidad, documentar la supervisión es también la mejor defensa jurídica disponible.

- **Las métricas de aceptación y derivación son de adopción inmediata.** No requieren compra de tecnología: requieren registrar lo que los revisores ya hacen. Una entidad colombiana podría publicarlas este año y convertirse en referente regional.
- **La matriz habilita, no solo restringe.** Clasificar decisiones por impacto y reversibilidad identifica también dónde la automatización plena es legítima —trámites masivos de bajo impacto—, liberando capacidad humana para los casos donde el juicio importa.
- **Ventana regulatoria.** Con el reglamento europeo aún sin aplicar y el proyecto brasileño en trámite, las entidades que adopten ahora estándares verificables influirán en cómo se escriban las reglas nacionales, en lugar de recibirlas.

Riesgos.

- **Supervisión teatral.** El riesgo principal es cumplir con la letra: designar revisores sin tiempo ni autoridad. La evidencia de sesgo de automatización indica que esa configuración produce resultados iguales o peores que la automatización declarada, con la ventaja adicional —para el sistema, no para el ciudadano— de un firmante humano a quien culpar.
- **Presunción de fiabilidad.** Horizon muestra el peligro de tratar la salida de un sistema como evidencia autosuficiente. Toda entidad debería poder responder: si el sistema y un ciudadano afirman cosas contrarias, ¿qué procedimiento decide?
- **Doble velocidad regional.** Si la regulación exigible llega a Europa y Brasil pero no al resto de la región, los proveedores podrían reservar las versiones auditables de sus sistemas para los mercados regulados.
- **Costo de oportunidad del péndulo.** Una falla visible de automatización sin supervisión puede llevar a prohibiciones generales que eliminen también los usos legítimos. La matriz existe para evitar ambos extremos.

Perspectivas.

Con la información disponible a febrero de 2024, cabe esperar: la adopción formal del Reglamento de IA europeo durante 2024, con la supervisión humana como obligación central para sistemas de alto riesgo; la publicación del memorando definitivo de la OMB en Estados Unidos; avances del PL 2338/2023 en Brasil; y la decisión de la Corte Constitucional colombiana sobre el caso de Cartagena, que probablemente fijará las primeras reglas jurisprudenciales de la región sobre IA generativa en la función judicial. No sabemos aún si alguna jurisdicción convertirá las métricas de supervisión —aceptación, derivación— en obligaciones de reporte; ninguna lo ha hecho a la fecha, y esa ausencia es la brecha que este documento propone cerrar por vía administrativa.

Recomendaciones.

Para entidades públicas que usan o planean usar sistemas de apoyo a decisiones:

1. **Clasifique cada sistema en la matriz antes de discutir tecnología.** Impacto sobre derechos y reversibilidad del error determinan el modo de supervisión; el proveedor no participa en esa clasificación.
2. **Excluya por escrito las seis categorías indelegables.** Un párrafo en el acto administrativo de adopción evita años de ambigüedad: qué decisiones de la entidad nunca serán salida de un sistema.
3. **Dote la supervisión con las cinco condiciones** —competencia, información, tiempo, autoridad, registro— y documéntelas. Si no puede financiar revisores con tiempo real, reduzca el alcance del sistema: la supervisión imposible de costear es señal de que la automatización propuesta excede la capacidad institucional.
4. **Mida y publique aceptación y derivación.** Trimestralmente, por sistema. Son dos números; su publicación convierte el principio de control humano en hecho verificable.
5. **Establezca la regla de contradicción.** Defina por procedimiento qué ocurre cuando el ciudadano y el sistema afirman cosas incompatibles: quién investiga, en qué plazo y con qué carga de la prueba. Horizon es lo que ocurre cuando esa regla no existe.
6. **Entrene a los revisores en las limitaciones del sistema**, no solo en su manejo: dominio de diseño, tasas de error conocidas, tipos de caso donde falla. Un revisor que no sabe cuándo desconfiar no puede supervisar.

Metodología.

Revisión de fuentes públicas con corte al 29 de febrero de 2024: informes oficiales de investigación (Comisión Real australiana, 2023; comité parlamentario neerlandés, 2020; Ofqual, 2020), decisiones judiciales (Tribunal de La Haya, 2020), normas e instrumentos (RGPD; texto acordado del Reglamento de IA; Directiva canadiense; Marco Ético colombiano) y literatura revisada por pares sobre sesgo de automatización e interacción humano-IA. Las cifras de los casos provienen de los documentos oficiales citados; donde las fuentes difieren (número exacto de afectados en Robodebt y Toeslagenaffaire), se usa la cifra más conservadora de fuente oficial.

Limitaciones. La lista de decisiones indelegables y la matriz de delegación son propuestas normativas de Enpath, derivadas del análisis jurídico y de la evidencia de fallas, no de validación experimental. Los casos examinados provienen de jurisdicciones con alta capacidad de investigación pública; es probable que fallas comparables en la región no estén documentadas con el mismo detalle, lo que sesga la muestra disponible hacia países de la OCDE.

Referencias.

1. Autoriteit Persoonsgegevens (2021). *Tax Administration fined for discriminatory and unlawful data processing*. Diciembre de 2021. https://www.autoriteitpersoonsgegevens.nl/en/current/tax-administration-fined-for-discriminatory-and-unlawful-data-processing
2. BID (2021). *PretorIA y la automatización del procesamiento de causas de derechos humanos*. Banco Interamericano de Desarrollo, fAIr LAC. https://publications.iadb.org/es/pretoria-y-la-automatizacion-del-procesamiento-de-causas-de-derechos-humanos
3. Buçinca, Z., Malaya, M. B. y Gajos, K. Z. (2021). *To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making*. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1). https://doi.org/10.1145/3449287
4. Consejo de la Unión Europea (2024). *Endoso del texto final del Reglamento de IA por el Comité de Representantes Permanentes*. 2 de febrero de 2024. https://www.consilium.europa.eu/en/policies/artificial-intelligence/
5. Gobierno de Canadá (2019, actualizada 2023). *Directive on Automated Decision-Making*. Treasury Board Secretariat. https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592
6. Gobierno del Reino Unido (2024). *New law to quash convictions of postmasters in Horizon scandal*. Enero de 2024. https://www.gov.uk/government/news/new-law-to-quash-convictions-of-postmasters-in-horizon-scandal
7. Green, B. y Chen, Y. (2019). *Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments*. Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* 2019). https://doi.org/10.1145/3287560.3287563
8. Ofqual (2020). *Summer 2020 grades for GCSE, AS and A level*. Agosto de 2020. https://www.gov.uk/government/news/summer-2020-grades-for-gcse-as-and-a-level
9. Parlamento de los Países Bajos (2020). *Ongekend onrecht* (Injusticia sin precedentes). Informe de la comisión parlamentaria sobre los subsidios de cuidado infantil. Diciembre de 2020. https://www.tweedekamer.nl/kamerstukken/detail?id=2020D50046
10. Presidencia de la República de Colombia (2021). *Marco Ético para la Inteligencia Artificial en Colombia*. https://dapre.presidencia.gov.co/TD/MARCO-ETICO-PARA-LA-INTELIGENCIA-ARTIFICIAL-EN-COLOMBIA-2021.pdf
11. Rechtbank Den Haag (2020). *Sentencia sobre el sistema SyRI*. ECLI:NL:RBDHA:2020:865. 5 de febrero de 2020. https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:865
12. Reglamento (UE) 2016/679 (RGPD), artículo 22. https://eur-lex.europa.eu/eli/reg/2016/679/oj
13. Royal Commission into the Robodebt Scheme (2023). *Report*. Gobierno de Australia, 7 de julio de 2023. https://robodebt.royalcommission.gov.au/publications/report
14. Skitka, L. J., Mosier, K. L. y Burdick, M. (1999). *Does automation bias decision-making?* International Journal of Human-Computer Studies, 51(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

Enpath · Investigación aplicada para decisiones públicas basadas en evidencia. Cada publicación incluye su metodología, sus fuentes y sus limitaciones.