AMEC EDICIONES

Privacidad y protección de datos en la era de la IA

Dr. Fernando Bonete Vizcaíno

Nuevas Tecnologías y Sociedad de la Información

RUTA DIRECTA A LA INNOVACIÓN DOCENTE

Privacidad y protección de datos en la era de la IA
Colección Ruta Directa a la Innovación Docente nº 96
2026 AMEC Ediciones Calle Emma Penella 6. 28055. Madrid. España.
ISBN: 978-84-10426-98-6
https://doi.org/10.63083/lamec.2026.14.fbv

Índice de contenidos

Situación de partida

- La irrupción de la inteligencia artificial a nivel usuario llega en 2018-2019 y para 2021 es arrolladora.
- Adopción masiva de las herramientas de IA a nivel usuario.
- Uso "personal" de la IA para tareas que enlazan con el trabajo en mi organización.
- Las organizaciones "tardan" en adquirir el uso de estas herramientas.
- En paralelo, la comisión Europea, viendo venir la marea, comienza a trabajar en su regulación con el Grupo Independiente de Expertos de Alto Nivel sobre Inteligencia Artificial hasta desembocar en la Reglamento de Inteligencia Artificial (RIA, julio de 2024).

¿Dónde está el análisis de riesgos?

- Robo de datos y pérdida de la confidencialidad o del anonimato.
- Inferencia de información.
- Falta de transparencia y de control de datos.
- Falta de calidad en los datos (sesgos, errores, incompletitud).
- Desvío de la finalidad: los datos se recogen para A, pero el uso es B.

Los resultados recabados por los principales estudios acerca de la preocupación por este tema revelan que entre el 80%–90% percibe la IA como una amenaza para la privacidad, entre un 75%–85% desconoce qué se hace con sus datos y quién tiene en realidad acceso a esos datos, pero...

Solo el 10% toman medidas o se interesan por conocer la situación a fondo.

La paradoja de la privacidad

Disonancia cognitiva

Center for Technology Innovation at BROOKINGS

JANUARY 2017

The privacy paradox II: Measuring the privacy benefits of privacy threats

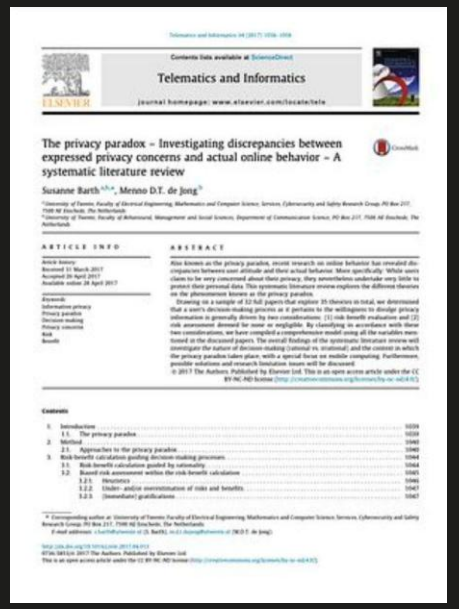
Benjamin Wittes and Emma Kohse

INTRODUCTION

100 THE JOURNAL OF CONSUMER AFFAIRS

PATRICIA A. NORBERG, DANIEL R. HORNE, AND DAVID A. HORNE

The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors



Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review

Susanne Barth, Menno D.T. de Jong

RGPD – RIA

- Principios rectores comunes.
- RPGD = Lex generalis
- RIA = Lex specialis
- Más allá de la RGPD, la AEPD pauta cómo incide una tecnología en la protección de datos
- La RIA también contempla riesgos para la salud, para la seguridad, para el resto de derechos fundamentales, para el medio ambiente.

Enfoque al riesgo: establecimiento de medidas en función del riesgo que entraña el uso de datos para la preservación de derechos y libertades.

A mayor riesgo, normas más estrictas

Riesgo mínimo o nulo

La mayoría de los sistemas de IA no plantean riesgos. Los juegos o los filtros de correo no deseado basados en la IA pueden utilizarse libremente, puesto que **no están regulados** ni se ven afectados por el Reglamento de IA de la UE.

Riesgo limitado

Los sistemas de IA que comportan un riesgo limitado, como los *chatbots* o los sistemas de IA generadores de contenido, están sujetos a **obligaciones de transparencia**, como la de informar a los usuarios de que su contenido se ha generado mediante IA para que puedan tomar decisiones con conocimiento de causa sobre su uso posterior.

Riesgo alto

Los sistemas de IA de alto riesgo, como los utilizados en el diagnóstico de enfermedades, la conducción autónoma y la identificación biométrica de personas implicadas en actividades delictivas u objeto de investigaciones penales, deben cumplir unos **requisitos y obligaciones estrictos** para acceder al mercado de la UE. Entre ellos se incluyen unas pruebas rigurosas, transparencia y supervisión humana.

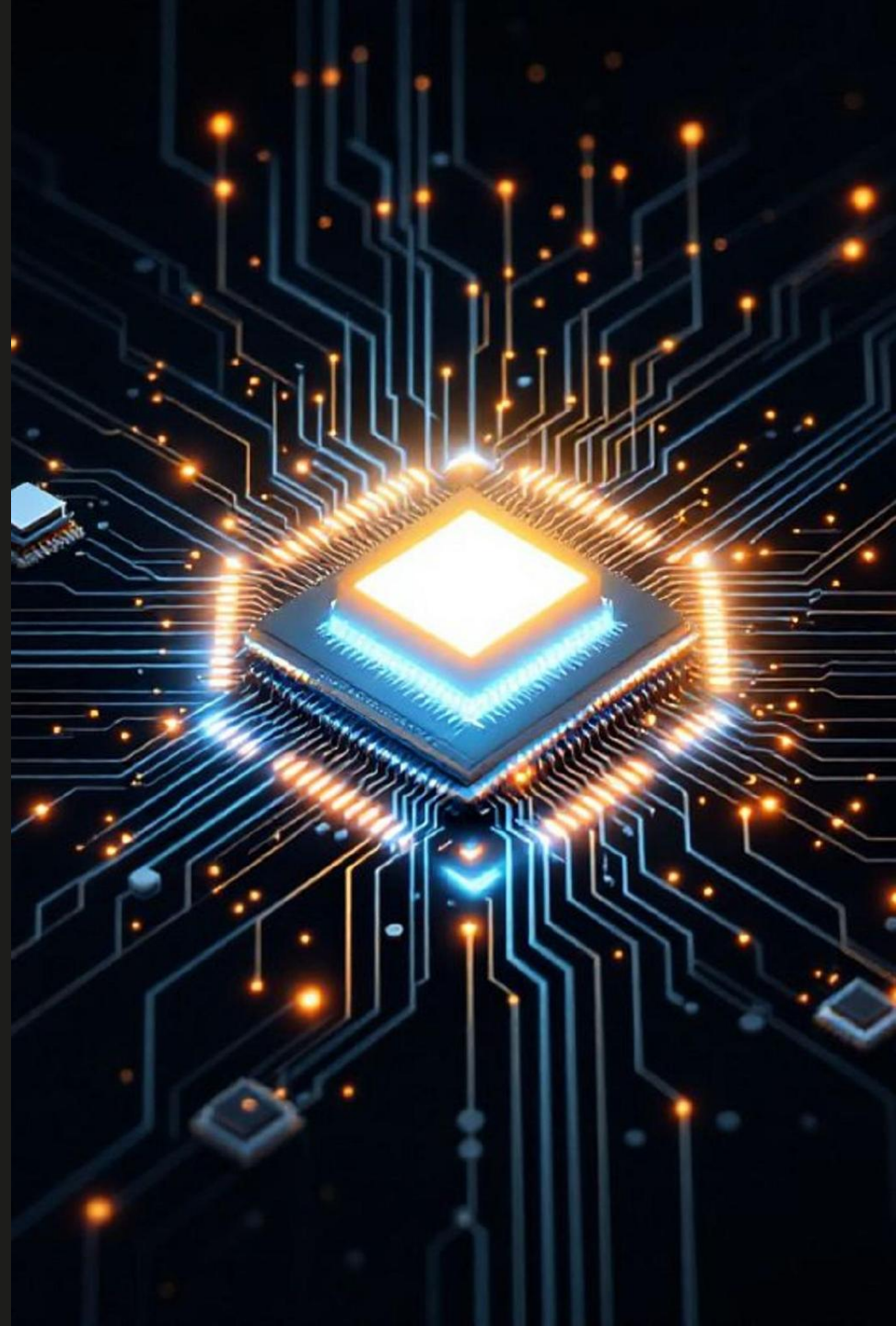
Riesgo inaceptable

Está prohibido en la UE el uso de sistemas de IA que supongan una amenaza para la seguridad, los derechos o los medios de subsistencia de las personas. Están prohibidos, por ejemplo, la manipulación cognitiva conductual, la actuación policial predictiva, el reconocimiento de emociones en lugares de trabajo y centros educativos, y la puntuación ciudadana, así como —con algunas excepciones— el uso en los espacios públicos, por parte de las autoridades garantes del cumplimiento del Derecho, de sistemas de identificación biométrica remota en tiempo real, como el reconocimiento facial.

Prohibiciones generales en el uso de IA

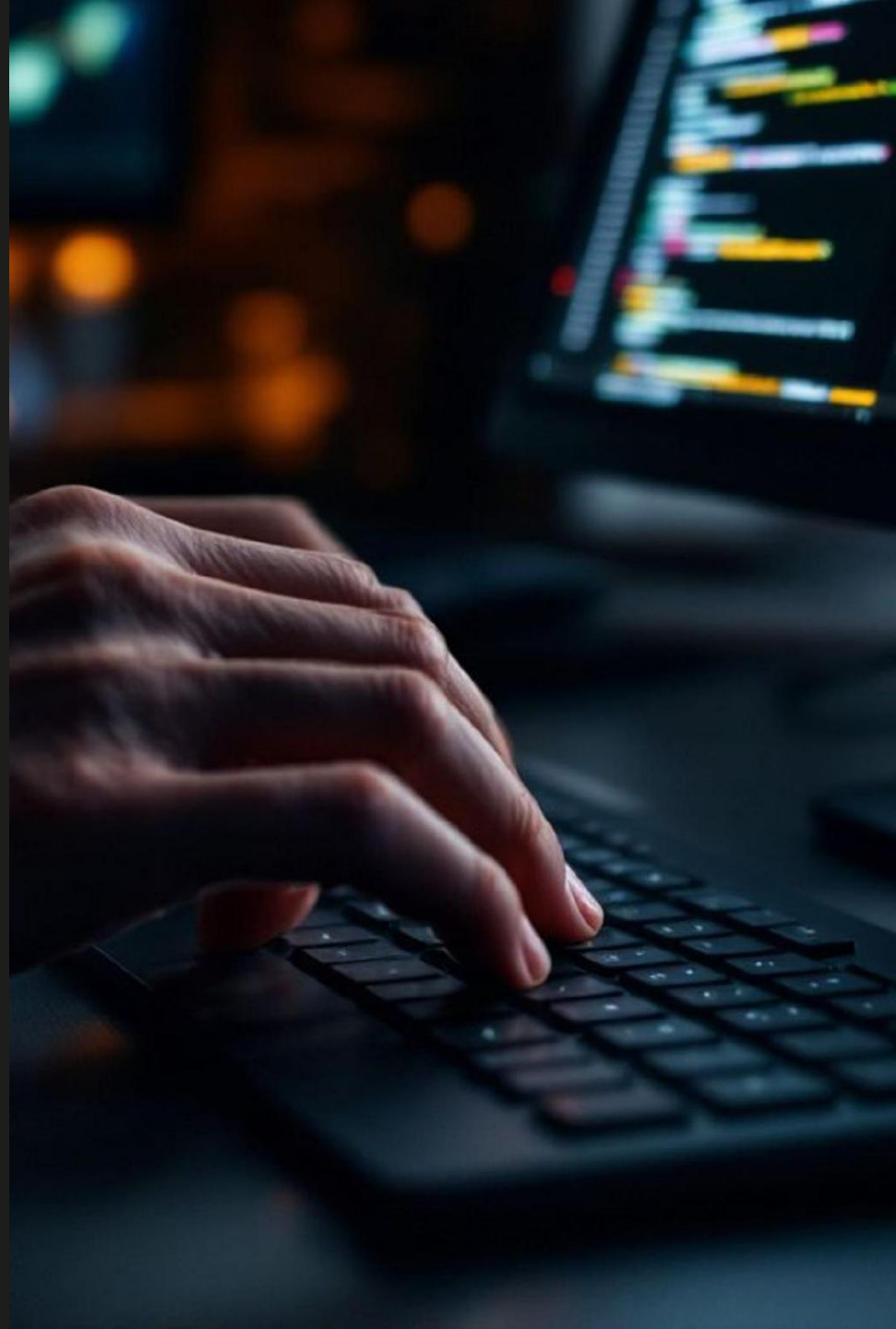
Prácticas que puedan causar manipulación, discriminación o invasión de la privacidad.

- **Manipulación subliminal o engañosa**: está prohibido el uso de IA que manipule el comportamiento de una persona sin su conocimiento.
- **Explotación de vulnerabilidades**: no se permite el uso de IA para explotar a personas en situaciones de vulnerabilidad (edad, discapacidad, pobreza).
- **Puntuación social basada en IA**: no se pueden evaluar o clasificar a las personas según su comportamiento social o características personales de manera discriminatoria.
- **Evaluación del riesgo delictivo basada en IA**: está prohibido predecir la probabilidad de que una persona cometa un delito basándose en su perfil o características personales.



Prohibiciones generales en el uso de IA

- **Reconocimiento facial masivo**: no se permite la creación de bases de datos de reconocimiento facial extrayendo imágenes de internet sin consentimiento.
- **Inferencia de emociones en el trabajo o educación**: está prohibido el uso de IA para analizar emociones en entornos laborales o educativos, salvo por razones médicas o de seguridad.
- **Categorización biométrica basada en datos sensibles**: no se puede utilizar IA para deducir raza, orientación política, afiliación sindical, creencias religiosas, vida sexual u orientación sexual a partir de datos biométricos.
- **Identificación biométrica en espacios públicos en tiempo real**: su uso con fines policiales solo está permitido en circunstancias muy específicas (búsqueda de personas desaparecidas, prevención de terrorismo, localización de sospechosos de delitos graves).



¿Qué pasa con la información que introducimos en un sistema general de IA?

Flujo de la Información en un Sistema de IA

Este diagrama ilustra cómo fluye la información a través de un sistema de Inteligencia Artificial, desde la entrada inicial hasta su posible uso futuro o eliminación.

1

Entrada de datos

Ingreso inicial de la información al sistema.

2

Procesamiento interno

Análisis y transformación de los datos.

3

Uso en el modelo

Aplicación de la información para generar resultados.

4

Almacenamiento

Guarda temporal o permanente de los datos.

5

Entrenamiento futuro

Posible uso para mejorar el modelo.

6

Compartir datos

Posible intercambio con terceros.

7

Eliminación o retención

Decisión sobre la permanencia de los datos.

Buenas prácticas en el uso de inteligencia artificial general para garantizar el derecho a la protección de datos

1. Cumplimiento normativo

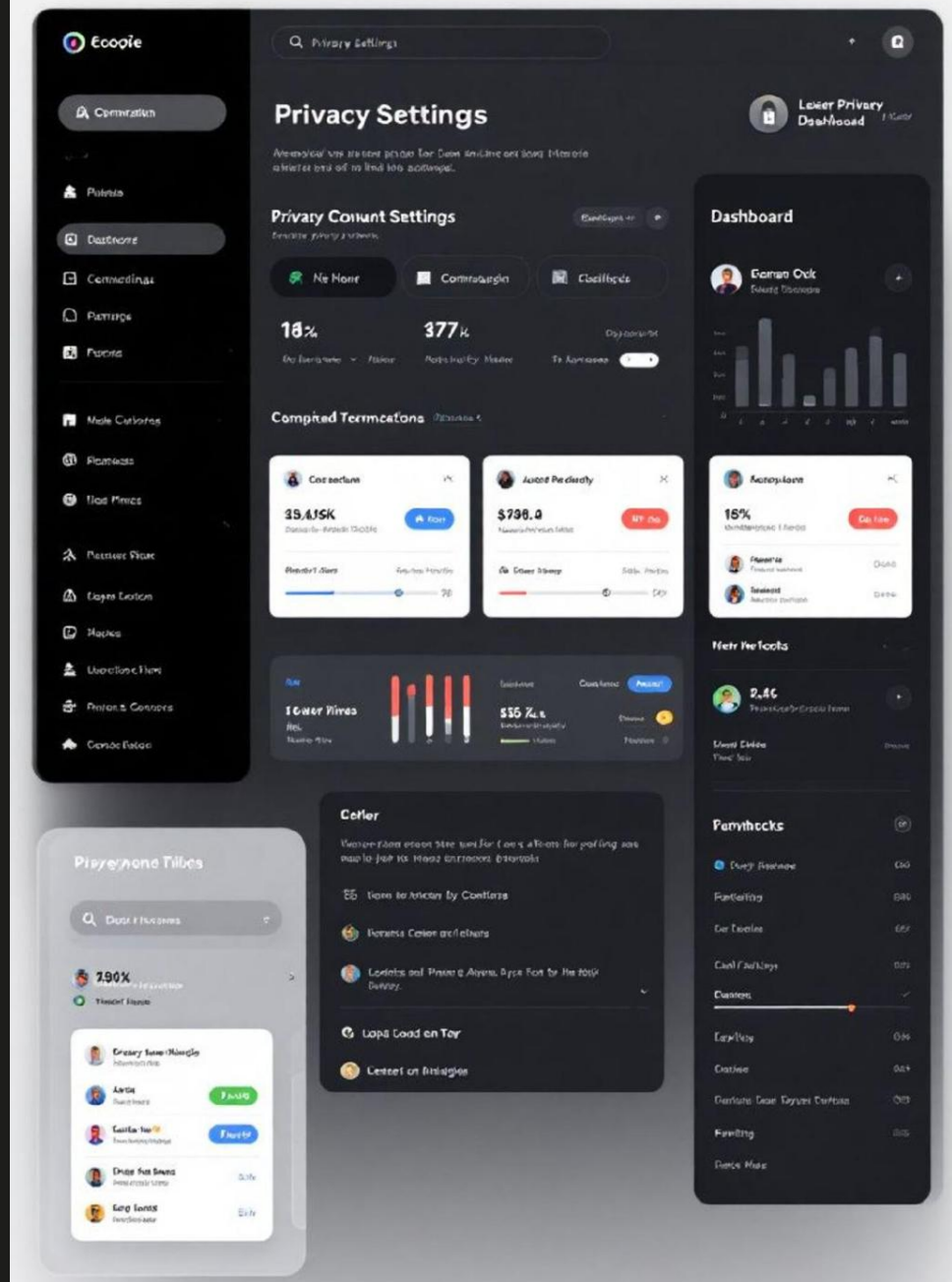
- Asegúrate de que las herramientas de IA que utilizas cumplen con normativas de protección de datos.
- Verifica que los proveedores de IA que utilizas cuentan con políticas de privacidad claras y mecanismos para gestionar la privacidad de los datos para las tareas a realizar en cada momento.

2. Minimización de datos

- Proporciona los datos estrictamente necesarios para el funcionamiento de la herramienta.
- Evita compartir información sensible o datos personales identificables cuando no sea imprescindible.
- Utiliza datos anonimizados o seudonimizados para reducir los riesgos de exposición.

3. Configuración de privacidad

- Revisa y ajusta las configuraciones de privacidad en cada plataforma de IA utilizada.
- Desactiva el almacenamiento de interacciones si la herramienta permite esta opción.
- Opta por servicios que permitan autogestión de datos y eliminación de información si es necesario.



4. Seguridad en la transmisión y almacenamiento de datos

- Utiliza herramientas que ofrezcan **cifrado de extremo a extremo** y protocolos seguros como **HTTPS, TLS o VPN** para proteger la comunicación de datos.
- Si almacenas datos procesados por IA, asegúrate de hacerlo en entornos seguros con acceso restringido.
- Implementa autenticación multifactor (MFA) en las cuentas vinculadas a plataformas de IA.

5. Transparencia y control

- Infórmate sobre cómo procesa y almacena la IA los datos que proporcionas.
- Verifica si los modelos de IA utilizados entrenan con los datos proporcionados o si se eliminan tras su procesamiento.
- Usa herramientas de auditoría para evaluar el tratamiento de datos en plataformas de IA.

6. Uso ético de la IA

- Evita el uso de IA para recopilar o analizar datos personales sin el consentimiento informado de los usuarios.
- Si trabajas con IA en un entorno empresarial, establece **políticas internas** sobre su uso responsable y la protección de datos.
- Asegura que cualquier toma de decisiones automatizada basada en IA sea transparente y explicable.

7. Evaluación de proveedores de IA

- Opta por herramientas de IA de proveedores reconocidos que cumplan con estándares de seguridad y privacidad.
- Lee las políticas de privacidad y los términos de uso para saber cómo gestionan los datos.
- Da preferencia a soluciones de IA **on-premise** o de código abierto si deseas mayor control sobre la información.

Fair play en el contexto de la IA

El concepto de *fair play* en la inteligencia artificial (IA) se refiere a la implementación de sistemas algorítmicos y modelos de aprendizaje automático que operen de manera equitativa, imparcial y sin favorecer injustamente a determinados grupos o individuos.

Equidad

Sistemas que operan sin favoritismos

Imparcialidad

Algoritmos que toman decisiones objetivas

Justicia

Modelos que respetan la igualdad entre grupos e individuos

Importancia de la equidad en la IA

Si una IA no es justa, puede reforzar y amplificar prejuicios existentes en la sociedad, perpetuando desigualdades y discriminaciones.

Los principales motivos por los que es esencial garantizar la equidad en la IA incluyen:

1 Justicia social y derechos humanos

garantizar la equidad en la IA ayuda a prevenir discriminaciones y promueve un acceso igualitario a los recursos y oportunidades.

2 Credibilidad y confianza

los escándalos de sesgo algorítmico han minado la confianza en la IA, generando preocupación sobre su impacto.

3 Cumplimiento normativo y regulaciones

empresas y organizaciones deben asegurarse de que sus modelos cumplen con normativas como el Reglamento de IA de la Unión Europea para evitar sanciones y problemas legales.

4 Impacto económico y reputacional

un caso paradigmático es el de Amazon, que en 2018 tuvo que desechar un sistema de reclutamiento basado en IA porque discriminaba sistemáticamente a las mujeres.

Casos de estudio

El sistema de reclutamiento de Amazon

Amazon desarrolló un sistema de inteligencia artificial para evaluar currículums y automatizar el proceso de selección de personal. La idea era reducir el tiempo y el esfuerzo necesario para filtrar y clasificar candidatos, confiando en un algoritmo que identificara a los postulantes más adecuados para cada puesto.

El sistema analizaba los CV de los aspirantes y los puntuaba con base en su adecuación al puesto, permitiendo que los reclutadores se enfocaran en aquellos que obtenían las mejores calificaciones.

Recepción de CV

El sistema recibía los currículums de todos los candidatos

Análisis automatizado

La IA analizaba el contenido de cada currículum

Puntuación

Asignaba una calificación según la adecuación al puesto

Filtrado

Los reclutadores recibían una lista ordenada por puntuación

El problema del sistema de Amazon

¿Y el problema?

En la práctica, el sistema:

Penalización de términos femeninos

Penalizaba los currículums que contenían la palabra "women" (mujeres), por ejemplo, en expresiones como "women's chess club" (club de ajedrez femenino) o "women's studies" (estudios de género).

Priorización masculina

Daba prioridad a candidatos con un lenguaje y experiencia similar a los CV de hombres contratados anteriormente.

Descarte automático

Descartaba automáticamente ciertos perfiles femeninos sin una justificación objetiva.

Lecciones del caso Amazon

El caso de Amazon dejó varias enseñanzas clave sobre la equidad en la IA:

Los datos de entrenamiento determinan el comportamiento del algoritmo

si los datos reflejan desigualdades previas, la IA las replicará.

La IA no es neutral

los algoritmos aprenden de patrones pasados y pueden amplificar prejuicios existentes si no se diseñan con un enfoque de equidad.

Es fundamental realizar auditorías constantes

los sistemas de IA deben ser revisados regularmente para detectar y corregir posibles sesgos antes de su implementación generalizada.

Los sesgos en IA no siempre pueden corregirse fácilmente

en algunos casos, el sesgo está tan arraigado en los datos de entrenamiento que la única solución viable es descartar el sistema.

El caso de COMPAS: sesgo en el sistema judicial

COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) es un software utilizado en Estados Unidos para evaluar la probabilidad de reincidencia de los acusados en procesos judiciales.

¿Cómo funciona COMPAS?

El software utiliza un algoritmo que analiza múltiples variables relacionadas con el historial delictivo, el contexto social y las respuestas de los acusados a cuestionarios de evaluación de riesgo. Con base en estos datos, COMPAS asigna a cada persona un puntaje de riesgo que indica la probabilidad de que reincida en el futuro.

¿Cuál era el problema?

El análisis de *ProPublica*, basado en más de 7,000 casos en Florida, encontró que:

- COMPAS **sobrestimaba el riesgo de reincidencia de los acusados afroamericanos**, asignándoles calificaciones de riesgo más altas en comparación con acusados blancos con antecedentes similares.
- Por el contrario, el sistema **subestimaba el riesgo de reincidencia de acusados blancos**, otorgándoles calificaciones más bajas incluso en casos donde tenían un historial delictivo más grave.

Sesgo en la publicidad algorítmica

Sesgo en la publicidad algorítmica: Discriminación de género en anuncios de empleo

Un estudio de la Universidad de Carnegie Mellon analizó los anuncios de Google y descubrió que **los anuncios de empleos mejor remunerados se mostraban con mayor frecuencia a usuarios masculinos que a femeninos**, incluso cuando ambos tenían perfiles similares en términos de búsqueda y experiencia laboral.

1

Aprendizaje de patrones históricos

Los algoritmos aprenden patrones a partir de los datos de clics históricos.

2

Refuerzo de tendencias

Si más hombres han hecho clic en anuncios de empleos bien pagados en el pasado, el sistema asume que los futuros clics provendrán de ellos y les muestra más estos anuncios.

3

Consecuencia discriminatoria

Como resultado, **las mujeres ven menos anuncios de empleos mejor remunerados**, lo que puede reforzar desigualdades de acceso laboral.

¿Qué son los sesgos computacionales?

Un **sesgo computacional** es una desviación sistemática en el comportamiento de un sistema de inteligencia artificial que conduce a resultados erróneos, desiguales o discriminatorios para ciertos grupos de usuarios.

Los sesgos en la IA **no surgen de la nada**:

Datos de entrenamiento sesgados

Información que contiene prejuicios históricos

Diseño del algoritmo

Decisiones técnicas que favorecen ciertos resultados

Errores de interpretación

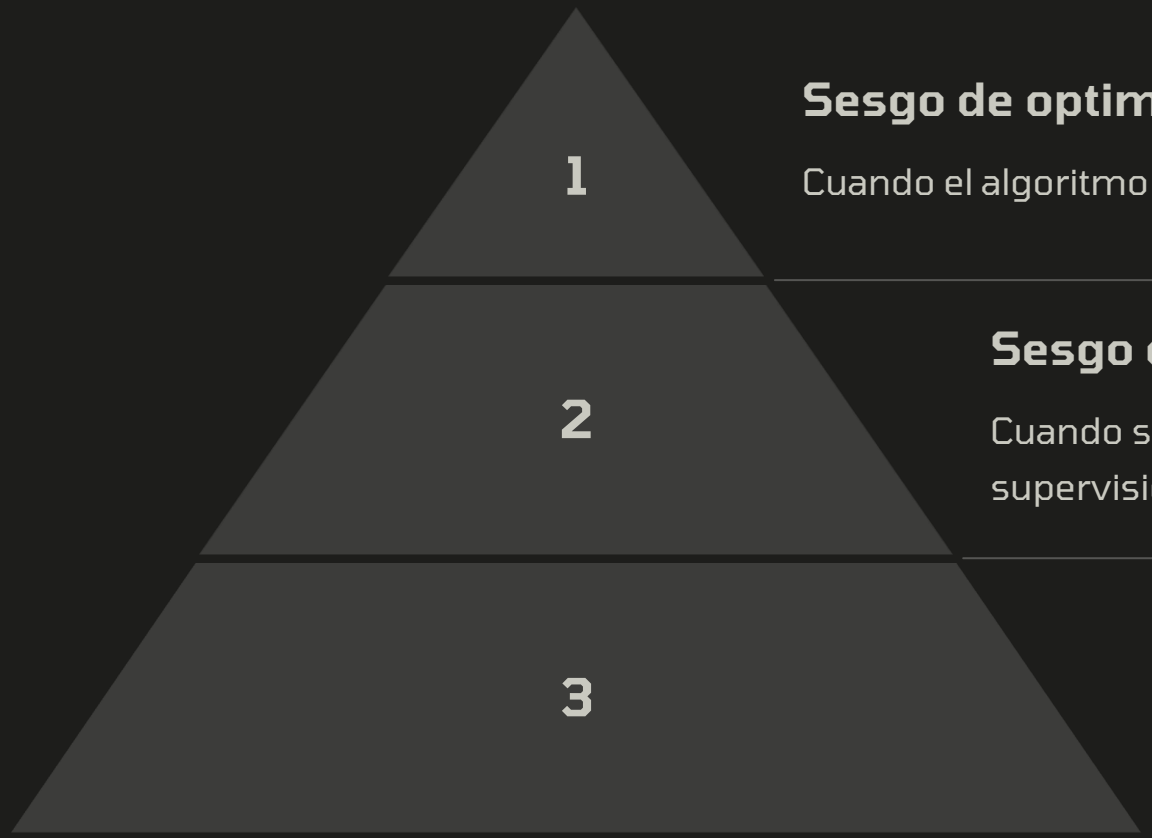
Malentendidos sobre los resultados del sistema

Falta de representatividad

Ausencia de diversidad en los datos

Tipos de sesgos computacionales
Sesgos algorítmicos

Los **sesgos algorítmicos** son aquellos que surgen del diseño, estructura y funcionamiento de los modelos de IA.



Sesgo de optimización

Cuando el algoritmo optimiza para un objetivo que genera resultados injustos

Sesgo de automatización

Cuando se confía excesivamente en sistemas automáticos sin supervisión

Sesgo de feedback loop

Cuando el sistema refuerza y amplifica patrones existentes

Sesgos de datos

Los **sesgos de datos** tienen lugar cuando la información utilizada para entrenar un sistema de IA no es representativa o contiene distorsiones que afectan las predicciones del modelo.

1

Sesgo de selección

Cuando la muestra de datos no representa adecuadamente a la población

2

Sesgo de recolección

Cuando el método de recopilación de datos introduce distorsiones

3

Sesgo de etiquetado

Cuando las categorías asignadas a los datos reflejan prejuicios

Métodos para detectar sesgos computacionales

Auditoría de datos y análisis de equidad

Auditoría de datos

La primera etapa en la detección de sesgos es la auditoría de los datos utilizados para entrenar los modelos de IA.

1. Análisis de representatividad.
2. Detección de desequilibrios.
3. Identificación de correlaciones no deseadas.
4. Evaluación de la fuente de datos.

Análisis de equidad en modelos de IA

Permite evaluar si los resultados de un sistema afectan de manera injusta a ciertos grupos de población.

- Comparación de tasas de aciertos y errores entre distintos grupos poblacionales.
- Evaluación del impacto desproporcionado en decisiones automatizadas.
- Análisis de fairness trade-offs.

Métodos estadísticos y métricas de equidad

1 Disparate Impact (impacto dispar)

mide si la tasa de selección de un grupo protegido es significativamente menor que la de un grupo privilegiado.

- Se calcula dividiendo la tasa de selección del grupo desfavorecido entre la tasa de selección del grupo privilegiado.
- **Ejemplo**: Si un modelo de contratación selecciona al 70% de los candidatos masculinos y al 40% de las mujeres, el disparate impact sería 0.57, lo que indica una desigualdad considerable.

2 Equalized Odds (probabilidades igualadas)

evalúa si la tasa de falsos positivos y falsos negativos es similar entre diferentes grupos.

Ejemplo: si un sistema de reconocimiento facial tiene una tasa de error del 5% para personas blancas, pero del 20% para personas negras, hay una clara violación de esta métrica.

3 Demographic Parity (paridad demográfica)

se refiere a la idea de que la proporción de personas seleccionadas por un modelo debe ser similar entre diferentes grupos.

4 Calibration (calibración del modelo)

comprueba si las predicciones del modelo son igualmente precisas para todos los grupos.

Estrategias para mitigar los sesgos en la IA

Recolección y curación de datos más representativos

- **Asegurar una representación equitativa** en los conjuntos de datos.
- **Recolectar datos de múltiples fuentes** para evitar sesgos inherentes a una sola perspectiva.
- **Eliminar datos irrelevantes o problemáticos** que puedan inducir sesgos.

Ajustes en el diseño de algoritmos y modelos

- **Reentrenamiento del modelo** con datos balanceados.
- **Ajuste de pesos en el algoritmo**.

Implementación de controles de equidad

- **Evaluaciones de impacto algorítmico** antes de lanzar un modelo.
- **Auditorías regulares** realizadas por equipos independientes.
- **Uso de explicabilidad en IA** para entender por qué un modelo toma ciertas decisiones.
- **Permitir la intervención humana** en decisiones críticas para evitar errores sistemáticos.

AMEC EDICIONES

¡Gracias!