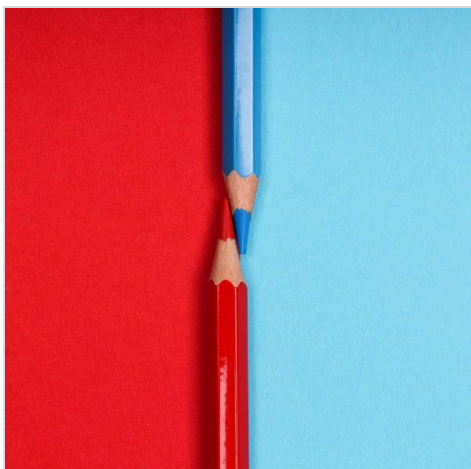


L U M B R E



Auditoría y evaluación de impacto de sistemas de IA

Grado: 8

Materia: Ética e IA Responsable · Formación Transversal

11/10/2026

Antes de empezar

Cómo pasar de los principios de IA responsable a evidencia: qué auditar, con qué artefactos (model cards, datasheets), cómo red teamear y por qué monitorear después.

QUÉ APRENDERÁS

- Diseñar una auditoría de un sistema de IA sobre sus datos, modelo y despliegue, apoyada en artefactos de trazabilidad y evaluación de impacto.

Auditoría y evaluación de impacto de sistemas de IA

VIDEO

Evaluación de impacto de sistemas de IA

Video explicativo (inglés): cómo se hace una evaluación de impacto algorítmico en el sector público (Government of Canada).



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=q6Pk2-UB8So>

Del principio a la evidencia

Un código de ética no encuentra el sesgo de tu modelo. Pasar de principios a práctica exige auditar el sistema concreto, documentarlo y vigilarlo después de desplegarlo.

Hablar de IA responsable es fácil; demostrar que un sistema concreto lo es, no. La **auditoría de IA** reúne evidencia sobre datos, modelo y despliegue; la **evaluación de impacto** estima daños posibles sobre personas y grupos antes de que ocurran.

Qué se audita

- Datos: representatividad, licencias, proveniencia y sesgos del corpus.

- Modelo: equidad entre grupos, robustez, calibración y fallos conocidos.
- Despliegue: monitoreo, límites de uso y control humano.
- Proceso: documentación, trazabilidad y responsabilidad.

Artefactos de trazabilidad

Documentar hace auditable. Los **model cards** reportan para qué se entrenó un modelo, con qué datos, su rendimiento por subgrupo y sus limitaciones. Las **datasheets** hacen lo mismo con un conjunto de datos, y los **system cards** describen un sistema completo, incluidos sus riesgos.



Auditar cruza los cuatro objetos con artefactos, tanto antes de publicar como después.

Evaluación de impacto

Antes de desplegar, se evalúa a quién afecta y cómo: poblaciones vulnerables, decisiones reversibles, magnitud del daño probable. La profundidad del análisis debe ser proporcional al riesgo del uso.

Red teaming y pruebas de robustez

- Adversariales: entradas diseñadas para romper el comportamiento.
- Equidad: medir disparidad de resultados entre grupos.
- Fuga: intentar extraer datos o instrucciones sensibles.

Monitoreo post-despliegue

Un modelo auditado una vez no queda auditable para siempre: los datos derivan y el contexto cambia. La vigilancia continua y un canal para reportar incidentes cierran el ciclo que el pre-despliegue abre.

Cada artefacto aporta evidencia de una capa distinta del sistema.

Artefacto	Qué registra
Datasheet	origen, composición y usos previstos de un dataset
Model card	rendimiento, límites y sesgos de un modelo
System card	comportamiento y riesgos de un sistema completo
Impact assessment	afectados y daños probables del despliegue

Errores frecuentes

- Auditar una métrica global y no mirar el desglose por subgrupos.
- Tratar la auditoría como un trámite de una sola vez, no un ciclo.
- Publicar un informe sin acciones: la evidencia exige cambios.

Ejemplo resuelto: auditar un modelo de selección de CV

1. Documentar el dataset con una datasheet: cómo se etiquetó y qué representa.
2. Medir rendimiento y disparidad por género y edad con un model card.
3. Evaluar impacto sobre solicitantes y definir límites de uso.
4. Hacer red teaming de sesgo y fugas antes de publicar.
5. Instalar monitoreo y un canal de reclamaciones tras el despliegue.

¿Para qué sirve en la realidad?

Con marcos legales de IA exigiendo documentación y gestión de riesgos, auditar deja de ser opcional. Un sistema capaz de mostrar su evidencia es defendible ante reguladores, clientes y ante su propio equipo.

1. ¿Qué distingue a un model card de una simple nota de versión?

- ☐ Que solo lista hiperparámetros
- ☐ Que documenta purpose, datos, rendimiento por subgrupo y limitaciones para hacer el modelo auditable
- ☐ Que reemplaza las pruebas de equidad
- ☐ Que garantiza que no hay sesgo

2. Auditar un sistema de IA es un evento único y no requiere vigilancia tras el despliegue.

☐ Verdadero

☐ Falso

3. El documento que registra el origen, composición y usos previstos de un conjunto de datos es una _____.

4. Clasifica cada artefacto según a qué objeto de auditoría documenta.

Elementos:

- procedencia y licencias del corpus
- rendimiento por subgrupo
- representatividad y sesgos de muestra
- límites y usos previstos del clasificador
- calibración y robustez

Datos	. . .
Modelo	. . .

5. Une cada término con su definición.

5.	Model card	-----
5.	Datasheet	-----
5.	Red teaming	-----
5.	Impact assessment	-----

Opciones: documenta propósito, datos y rendimiento del modelo · documenta el conjunto de datos · ataque controlado para hallar fallos · estima afectados y daños

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

6. Auditoría de IA



Auditoría y evaluación de impacto de IA

Auditar IA

Objetos

datos, modelo

despliegue, proceso

Artefactos

datasheet

model card

system card

Ciclo

evaluación de impacto

red teaming

monitoreo posterior

El mapa mental interactivo es interactivo en la versión online.



Model Cards for Model Reporting (Mitchell et al.) <https://arxiv.org/abs/1803.08382> Paper fundacional del artefacto model card para reportar modelos.

Datasheets for Datasets (Gebru et al.) <https://arxiv.org/abs/1803.09010> Trabajo que introduce la documentación sistemática de conjuntos de datos.

7. Un hospital quiere auditar un modelo de triaje antes de usarlo. ¿Qué artefactos de trazabilidad exigirías y cómo adaptarías el nivel de documentación al riesgo de dañar a pacientes?

Referencia curricular

Grado:	8
Materia:	Ética e IA Responsable · Formación Transversal

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-trans-etic-005-auditoria-evaluacion-de-impacto-de-ia>

Material educativo generado desde Lumbre · lumbre.lat