

L U M B R E



# Ética y datos desde el inicio

**Grado:** 1

**Materia:** Fundamentos de Ciencia de Datos · Análisis de Datos

11/10/2026

## Antes de empezar

La ética no es un trámite que se añade al final: se incorpora al concebir el proyecto. Comprender el origen de los datos, los sesgos que arrastran y el daño que un modelo puede causar es parte de la competencia profesional del científico de datos.

### QUÉ APRENDERÁS

- Reconocer las dimensiones éticas de un proyecto de datos —privacidad, sesgo, transparencia y rendición de cuentas— e integrarlas desde el diseño.

## Ética y datos desde el inicio

### Correcto no es lo mismo que justo

Un modelo puede ser técnicamente impecable y socialmente dañino. Predice bien si reproduce discriminaciones del pasado con aire de objetividad. La ética no frena la buena ciencia: la define.

Cada decisión del proyecto tiene una dimensión ética: qué datos se recogen y de quién, con qué consentimiento, qué se infiere sobre las personas y qué consecuencias tiene automatizar una decisión. Estas cuestiones no son accesorias; condicionan el diseño desde la primera pregunta.

### VIDEO

#### Ética y sesgo en ciencia de datos

Reconocer y mitigar los sesgos y los dilemas éticos del dato.



Escanea para verlo en Lumbre

[https://www.youtube.com/watch?v=C3\\_X1X9RRM4](https://www.youtube.com/watch?v=C3_X1X9RRM4)

## Origen y consentimiento de los datos

- ¿Los datos se obtuvieron con conocimiento y permiso válido?

- ¿Se usan para lo que se recogieron o para algo distinto?
- Ignorar el origen puede hacer ilegítimo un análisis técnicamente impecable.

## Privacidad y minimización

Se recogen solo los datos necesarios para la finalidad declarada. Cuanto menos se guarda, menos hay que proteger y menor es el daño si algo falla. La minimización es la primera defensa, no el cifrado.

## Sesgo en los datos

Un conjunto de datos hereda los prejuicios del mundo que describe. Si históricamente se contrató de cierta manera, un modelo que aprenda de esos datos no descubre mérito: replica discriminación y la disfraz de matemática.

## Sesgo en el modelo

- Las decisiones de modelado también cargan valores: qué métrica optimizar, a quién perjudica el error.
- Un umbral elegido puede excluir sistemáticamente a un grupo.
- Medir el rendimiento por subgrupos revela daños que el promedio esconde.



*Correcto no es lo mismo que justo: el poder técnico debe pesar sus efectos.*

## Transparencia y explicabilidad

Quien se ve afectado por una decisión automatizada suele tener derecho a una explicación. No todos los modelos son igual de interpretables, y elegir un modelo opaco para una decisión sobre personas exige una justificación fuerte.

## Responsabilidad y rendición de cuentas

- Alguien debe responder por el efecto del modelo; la responsabilidad no se delega en el algoritmo.
- Rastreabilidad de decisiones para poder revisarlas.
- Sin dueño claro, los daños caen en un vacío institucional.

## Finalidad y proporcionalidad

No todo lo que se puede predecir se debe predecir. Hay que preguntarse para qué se usará el resultado y si el beneficio justifica el riesgo sobre las personas. La proporcionalidad pone el poder técnico al servicio de un fin legítimo.

## Automatizar no es neutral

- Automatizar a escala amplifica tanto el acierto como el error.
- Retirar el juicio humano traslada poder sin asignar responsabilidad.
- Muchos sistemas necesitan un humano con capacidad de corregir.

## Regulación y derechos

Marcos legales protegen datos personales y, cada vez más, decisiones automatizadas. Conocerlos no es opcional: incumplir puede invalidar el proyecto completo y dañar a la organización.

## Ética como práctica, no como discurso

- Revisar el daño posible en cada fase, no solo al final.
- Incluir miradas diversas en el equipo que diseña.
- Dejar por escrito las decisiones y sus razones.

## Hacia una cultura responsable

La ética robusta es consecuencia de equipos que preguntan «deberíamos?» antes que «podemos?», y de procesos que premian detectar el problema temprano en vez de esconderlo.

## Errores frecuentes

- Tratar la ética como un trámite final antes de desplegar.
- Confiar en la aparente neutralidad del número para no mirar el daño.
- No designar un responsable cuando el modelo se equivoca.

## Ejemplo resuelto: revisar éticamente un modelo de contratación

Una empresa entrena un filtro de currículos con las contrataciones de los últimos años. El repaso ético:

1. Origen: comprobar que los datos de personal se usaron con consentimiento.
2. Sesgo: detectar si el histórico discrimina por género o procedencia.
3. Métrica: elegir una que no perjudique sistemáticamente a un grupo.
4. Explicabilidad: poder decir a un candidato por qué fue descartado.
5. Responsabilidad: designar quién responde y cómo se corrige el sistema.

## ¿Para qué sirve en la realidad?

Integrar la ética desde el diseño evita desplegar discriminaciones con aire de objetividad y protege a las personas y a la organización de daños y sanciones.

### 1. Un modelo de selección de personal entrenado con decisiones de contratación del pasado tiende a:

- ☐ eliminar todo sesgo por ser objetivo
- ☐ reproducir las discriminaciones históricas como si fueran mérito
- ☐ garantizar igualdad automáticamente
- ☐ ser siempre ilegal

### 2. La consideración ética de un proyecto de datos puede dejarse para la fase final, justo antes de desplegar.

- ☐ Verdadero
- ☐ Falso

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

### 3. Ética



#### 4. Clasifica cada situación según el principio ético que pone a prueba.

##### Elementos:

- Vender listados sin el permiso de las personas
- El modelo rechaza más a un grupo por su código postal
- Un ciudadano no entiende por qué le negaron el crédito
- Nadie en la empresa responde cuando el modelo falla
- Guardar datos «por si acaso» durante años
- Optimizar una métrica que perjudica a una minoría

|                        |       |
|------------------------|-------|
| <b>Privacidad</b>      | . . . |
| <b>Sesgo</b>           | . . . |
| <b>Transparencia</b>   | . . . |
| <b>Responsabilidad</b> | . . . |

Ética de la IA [https://en.wikipedia.org/wiki/Ethics\\_of\\_artificial\\_intelligence](https://en.wikipedia.org/wiki/Ethics_of_artificial_intelligence) Debate sobre impacto, sesgo y responsabilidad de los sistemas automáticos.

#### 5. Une cada principio ético con su idea clave:

|    |                 |       |
|----|-----------------|-------|
| 5. | Minimización    | ----- |
| 5. | Justicia        | ----- |
| 5. | Explicabilidad  | ----- |
| 5. | Responsabilidad | ----- |

**Opciones:** recolectar solo los datos necesarios · no reproducir ni amplificar sesgos · poder rendir cuentas de una decisión · siempre hay un humano que responde

#### Diseñar con criterio

Ética del dato

Origen

consentimiento válido

finalidad declarada

Riesgos

sesgo heredado

daño al automatizar

Deberes

privacidad y minimización

transparencia

## Cuentas

responsable claro

decisiones rastreables

El mapa mental interactivo es interactivo en la versión online.



Sesgo algorítmico [https://en.wikipedia.org/wiki/Algorithmic\\_bias](https://en.wikipedia.org/wiki/Algorithmic_bias) Errores sistemáticos que producen resultados injustos o discriminatorios.

**6. Tu empresa puede predecir, con alta exactitud, qué empleados tienen riesgo de quedarse embarazada para «organizar turnos». Argumenta con dos principios éticos por qué no debería hacerlo, aunque sea técnicamente posible y legalmente borroso.**

## Referencia curricular

|                 |                                                     |
|-----------------|-----------------------------------------------------|
| <b>Grado:</b>   | 1                                                   |
| <b>Materia:</b> | Fundamentos de Ciencia de Datos · Análisis de Datos |

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



<https://univ.lumbre.lat/object/cdia-fcd-008-etica-y-datos-desde-el-inicio>

Material educativo generado desde Lumbre · [lumbre.lat](https://lumbre.lat)