Actividad 3: Laboratorio de sesgos: ¿Es justo nuestro modelo?

Actividad 3. **¡Crea una IA justa! El laboratorio de datos y sesgos**

Justificación

Esta actividad hace visible que “aprender” en IA significa **crear un modelo** a partir de **datos** y luego **probarlo** con casos nuevos. El alumnado detecta **sesgos** cuando ciertos grupos quedan poco representados y entiende por qué eso puede ser injusto. Esta actividad Integra pensamiento crítico, convivencia y lenguaje de IA sin usar pantallas.

Datos de la actividad

- **Curso:** 5º Primaria (10–11 años)
- **Tiempo:** 55–70 min (RECOMENDADO EN 2 SESIONES)
- **Agrupación:** equipos de 4–5 + puesta en común
- **Espacio:** aula (mesas)

Objetivo general

Copyright 2026

Construir y evaluar un modelo sencillo (reglas/puntuación) para tomar una decisión, usando datos de ejemplo, y comprobar cómo aparece el **sesgo** por representación y por “reglas mal diseñadas”.

Materiales

- 30 “fichas de caso” (tarjetas) con Robots y rasgos neutros del aula (sin datos sensibles):
  - Ej.: “Trae deberes 4/5 días”, “Ayuda a un compañero”, “Se distrae a veces”, “Llega puntual”, “Interrumpe”, “Trabaja en equipo”…
- 3 sobres: **ENTRENAMIENTO (datos)** / **PRUEBA** / **DIFÍCILES**
- Plantilla “Nuestro modelo” (tabla de puntos o reglas)
- Hoja de registro: 12 predicciones □ /□
- Post-its: **DATOS** / **ALGORITMO** / **MODELO** / **SESGO** / **PRUEBA**

> Nota: Los “personajes” pueden ser animales o robots para hacerlo más lúdico.

Desarrollo de la actividad

1) Pregunta detonante (4 min)

“Si quisiéramos una ‘máquina’ que eligiera a quién elegir como el **Robot ayudante del aula**… ¿cómo le enseñaríamos a decidir?”

2) Organización (3 min)

Roles:

- **Diseñador/a del modelo** (escribe reglas/puntos)
- **Entrenador/a de datos** (analiza ejemplos)
- **Probador/a** (aplica el modelo a nuevos casos)
- **Auditor/a de sesgo** (revisa a quién perjudica y por qué)
- (Opcional) **Portavoz**

3) Datos (entrenamiento) (10–12 min)

Cada equipo recibe 10 tarjetas del sobre **ENTRENAMIENTO**, ya etiquetadas por el docente:

Copyright 2026

- “SÍ es elegido”
- “NO es elegido”

El equipo analiza:

- ¿Qué rasgos aparecen mucho en los “SÍ”?
- ¿Qué rasgos aparecen mucho en los “NO”?

> Se nombra: **DATOS = ejemplos con respuesta correcta**.

4) Construcción del algoritmo (modelo por puntos) (12–15 min)

El equipo crea un **algoritmo** tipo “puntos” (muy manipulativo y fácil de aplicar):

- +2 si “ayuda a compañeros”
- +2 si “respeta turnos”
- +1 si “trae material”
- −2 si “interrumpe”
- −1 si “llega tarde a veces”
  Regla final: si suma **≥ 3 puntos → SÍ**, si no → NO.

> Se nombra: ALGORITMO **= las reglas/pasos para sumar y decidir** y **MODELO = la tabla final de puntos (lo aprendido con los datos)**.

5) Prueba (validación) con datos nuevos (10–12 min)

Se entrega el sobre **PRUEBA** con 8 tarjetas nuevas (sin etiqueta).
El equipo aplica su modelo, predice SÍ/NO y registra.

Después el docente revela la etiqueta real (o una hoja de soluciones común).

6) Mini-laboratorio de sesgos (12–15 min)

Ahora entra el sobre **DIFÍCILES** con 4 tarjetas pensadas para mostrar sesgo. Ejemplos:

Copyright 2026

- Casos con rasgos “mixtos” que no aparecían en entrenamiento (“ayuda mucho pero se pone nervioso y habla rápido”)
- Casos con un rasgo que el modelo penaliza demasiado (“llega tarde a veces” por motivo ajeno, sin entrar en realidades personales)

El auditor responde en la hoja:

- ¿Qué tipo de caso recibe más “NO”?
- ¿Ese rasgo estaba poco representado en los datos?
- ¿La regla está exagerando una pista?

> Se nombra: **SESGO = cuando el modelo se equivoca más con ciertos casos porque los datos eran pobres o las reglas no eran justas.**

7) Mejora del modelo (6–8 min)

Cada equipo elige UNA mejora (obligatorio justificarla):

- **Mejora de datos:** añadir 2 tarjetas de entrenamiento que faltaban (más variedad de casos)
- **Mejora de reglas:** bajar un castigo (−2 → −1), subir un criterio positivo, o añadir “NO SÉ” si está en zona gris
- **Mejora de prueba:** crear una regla de “revisión humana” para casos límite

Se repiten las 4 tarjetas DIFÍCILES y se comparan resultados.

8) Debate final (5–7 min)

- “¿Qué mejoró más: cambiar datos o cambiar reglas?”
- “¿Cuándo sería peligroso usar este modelo sin revisar?”
- “¿Qué significa que un modelo sea ‘justo’?”

DUA (adaptaciones)

- **Representación:** tarjetas con iconos (mano ayudando, reloj puntual, boca hablando), lectura compartida.
- **Acción/expresión:** cálculo en grupo con fichas/puntos, explicación oral o con esquema.
- **Implicación:** rol de auditor (misión), debate con ejemplos concretos, opción “NO SÉ”.

Copyright 2026

Evaluación final

Semáforo por equipo

- □□ “Definimos datos/algoritmo/modelo/sesgo y mejoramos el modelo con una razón.”
- □□ “Lo entendemos pero nos cuesta explicar el sesgo o justificar la mejora.”
- □□ “Necesitamos apoyo para construir reglas y ver el sesgo.”

Checklist docente (5 ítems)

- Identifica datos como ejemplos etiquetados.
- Describe algoritmo como reglas/pasos.
- Explica modelo como la tabla final.
- Detecta sesgo (a quién perjudica y por qué).
- Propone una mejora razonada (datos o reglas).

Revision #4
Created 2026-01-16 00:05:42 CET by David Cañete
Updated 2026-01-20 17:51:31 CET by David Cañete

Copyright 2026