

Actividades Detalladas de la UT02: Modelos de Resolución de Problemas

El objetivo es que profundicemos no solo en la tecnología, sino también en sus implicaciones éticas y legales.

1. Debate Inicial: "¿Debe una IA tomar decisiones críticas sin supervisión humana?"

- **Objetivo:** Fomentar el pensamiento crítico sobre los límites de la autonomía en sistemas de IA y explorar las implicaciones éticas y de responsabilidad.
- **Formato:** Debate grupal moderado.
- **Enunciado:**
- Imaginad un futuro cercano donde los sistemas de IA son responsables de tomar decisiones críticas en tiempo real y sin posibilidad de consulta humana. Pensemos en tres escenarios concretos:
 1. Un **vehículo autónomo** que debe decidir entre atropellar a un peatón que ha cruzado indebidamente o desviarse bruscamente, poniendo en riesgo la vida de su ocupante.
 2. Un **sistema de diagnóstico médico** que, basándose en millones de casos, decide desconectar el soporte vital de un paciente con una probabilidad de supervivencia inferior al 0.1%, para asignar esos recursos a otro paciente con mayores posibilidades.
 3. Un **sistema de defensa militar autónomo** que detecta una amenaza inminente y debe decidir si lanza un contraataque que podría causar víctimas civiles colaterales.
- Para el debate, la clase se dividirá en dos grandes grupos: "A favor" y "En contra" de permitir que la IA tome estas decisiones de forma autónoma. Cada grupo deberá preparar argumentos sólidos, considerando aspectos como la **eficiencia**, la **eliminación del error humano**, la **responsabilidad legal** (¿quién es el culpable si la IA se equivoca?), los **sesgos algorítmicos** y los **dilemas éticos** fundamentales.

2. Análisis de un Caso Real: El Fracaso del Sistema de Contratación de Amazon

- **Objetivo:** Analizar un caso real para comprender cómo los sesgos en los datos de entrenamiento pueden llevar a resultados discriminatorios y cómo esto viola principios éticos y legales.
- **Formato:** Investigación individual o en parejas y posterior puesta en común.
- **Enunciado:**
- En 2018, la agencia de noticias Reuters desveló que Amazon había desechado un sistema de IA para reclutamiento porque mostraba un sesgo sistemático contra las mujeres. El modelo fue entrenado con los currículums recibidos por la compañía durante un período de 10 años, que en su mayoría pertenecían a hombres, lo que llevó a la IA a penalizar currículums que incluían la palabra "mujer" (como "capitana del equipo de ajedrez de mujeres") y a degradar a las graduadas de dos universidades exclusivas para mujeres.

- Vuestra tarea es investigar este caso y elaborar un informe de una página que responda a las siguientes preguntas:
 - **¿Cuál fue la causa raíz del sesgo en el algoritmo?** Explica cómo los datos históricos perpetuaron la discriminación.
 - **¿Qué principios éticos fundamentales se vulneraron?** (Equidad, justicia, no discriminación).
 - Si este sistema se hubiera implementado en la Unión Europea, **¿qué artículos del Reglamento General de Protección de Datos (RGPD) o de la futura Ley de Inteligencia Artificial (AI Act) podría haber infringido?**
 - **¿Qué medidas podría haber tomado Amazon** para mitigar o evitar este sesgo durante el diseño y entrenamiento del modelo?

3. Ejercicio Práctico: Identificando Riesgos y Sesgos en un Modelo de Concesión de Créditos

- **Objetivo:** Aplicar el pensamiento analítico para identificar vulnerabilidades éticas y operativas en un modelo de IA antes de su implementación.
- **Formato:** Ejercicio práctico individual.
- **Enunciado:**
- Un banco ha desarrollado un modelo de aprendizaje automático para decidir automáticamente si concede o no un préstamo personal a sus clientes. El modelo ha sido entrenado con datos de los últimos 20 años y utiliza las siguientes variables para tomar su decisión:
 - Edad del solicitante.
 - Código postal de residencia.
 - Nivel de ingresos anuales.
 - Historial crediticio (si ha tenido impagos).
 - Nivel de estudios.
 - Antigüedad en su puesto de trabajo actual.
- Tu misión es actuar como un auditor de IA. Analiza el planteamiento y **elabora una lista detallada de posibles riesgos y sesgos** que podría presentar este modelo. Para cada riesgo identificado, explica **por qué es un problema y qué colectivo podría verse afectado negativamente**. Piensa en sesgos evidentes y también en aquellos más sutiles (por ejemplo, ¿qué información indirecta puede ocultar un código postal?).

4. Role Play: Comité de Ética para "Health-AI"

- **Objetivo:** Simular un entorno profesional multidisciplinar para evaluar de forma integral la viabilidad ética y social de un proyecto de IA.
- **Formato:** Actividad grupal de simulación.
- **Enunciado:**
- Una startup tecnológica, "Health-AI", quiere lanzar una nueva aplicación móvil que utiliza IA para predecir la probabilidad de desarrollar enfermedades cardíacas en los próximos 5 años. La app recopilará datos de pulseras de

actividad (pasos, frecuencia cardíaca), un cuestionario de hábitos (dieta, tabaquismo) y datos demográficos del usuario.

- Formaréis equipos de 4 personas para simular el **Comité de Ética** de la empresa. Cada miembro asumirá un rol:
 1. **Jefe/a de Producto:** Defiende el proyecto, centrándose en su utilidad, modelo de negocio y beneficios para el usuario.
 2. **Ingeniero/a de IA:** Explica los aspectos técnicos, los datos necesarios y las posibles limitaciones del modelo.
 3. **Abogado/a especialista en protección de datos:** Se asegura de que el proyecto cumple con el RGPD y otras normativas. Su principal preocupación es la privacidad y el uso de datos sensibles.
 4. **Especialista en Ética y Sociedad:** Evalúa el impacto social, los posibles sesgos (¿funcionará igual para todos los grupos étnicos?), y el riesgo de generar ansiedad o discriminación (por ejemplo, por parte de aseguradoras).
- El objetivo del comité es **redactar un veredicto final** de una página: **aprobar el proyecto tal cual, aprobarlo con condiciones, o rechazarlo**, justificando la decisión basándose en los argumentos de todos los roles.

5. *Diseño de un Protocolo Básico de "Privacy by Design"*

- **Objetivo:** Traducir los principios teóricos de la privacidad desde el diseño a un conjunto de acciones y directrices prácticas para un proyecto concreto.
- **Formato:** Actividad individual o en parejas.
- **Enunciado:**
- Vais a diseñar el esqueleto de un protocolo de "Privacidad desde el Diseño" (*Privacy by Design*) para un proyecto ficticio: un **sistema de cámaras inteligentes para un comercio minorista**. Estas cámaras utilizan IA para analizar el comportamiento de los clientes: mapas de calor de las zonas más visitadas, tiempo medio que pasan en cada sección y perfil demográfico estimado (rango de edad y género) para optimizar la distribución de los productos.
- Vuestro protocolo debe incluir, al menos, los siguientes puntos en forma de directrices claras y concisas:
 - **Minimización de datos:** ¿Qué datos son estrictamente necesarios? ¿Cómo evitar la recopilación de información superflua?
 - **Anonimización y seudonimización:** ¿Qué técnicas se aplicarán a las imágenes para proteger la identidad de las personas *antes* de ser procesadas?
 - **Seguridad de los datos:** Medidas básicas para proteger los datos recopilados (cifrado, control de acceso).
 - **Transparencia y control del usuario:** ¿Cómo se informará a los clientes de que están siendo analizados? ¿Tienen alguna opción de oponerse?
 - **Política de retención:** ¿Durante cuánto tiempo se guardarán los datos y por qué?

6. *Elaboración de un Mapa Mental sobre el Marco Legal Aplicable*

- **Objetivo:** Estructurar y visualizar el complejo marco normativo que afecta a un caso de uso de IA, identificando las leyes y principios clave.
- **Formato:** Actividad individual con herramienta de mapas mentales (digital o manual).
- **Enunciado:**
 - Una empresa de logística quiere implementar un sistema de IA para monitorizar a sus repartidores. El sistema analizará en tiempo real los datos del GPS del vehículo, los tiempos de parada, la velocidad media y el cumplimiento de las rutas planificadas para evaluar la eficiencia y el rendimiento de cada empleado. Las evaluaciones generadas por la IA se usarán en las revisiones de desempeño anuales.
 - Crea un **mapa mental** que resuma el marco legal y ético que esta empresa debe tener en cuenta antes de lanzar el proyecto. El nodo central será "Sistema de Monitorización de Repartidores". A partir de ahí, crea ramas para los siguientes conceptos y desarróllalos con ideas clave:
 - **Reglamento General de Protección de Datos (RGPD):** Base legal para el tratamiento de datos, derechos de los trabajadores (acceso, rectificación, oposición), necesidad de una Evaluación de Impacto.
 - **Ley de IA de la UE (AI Act):** ¿Bajo qué categoría de riesgo caería este sistema (alto riesgo por afectar al ámbito laboral)? ¿Qué obligaciones implicaría?
 - **Derechos de los Trabajadores:** Estatuto de los Trabajadores, derecho a la intimidad y a la desconexión digital.
 - **Principios éticos:** Proporcionalidad (¿es necesaria tanta vigilancia?), transparencia (¿saben los empleados exactamente qué se mide y cómo?), y justicia algorítmica.