

ÉTICA EN LA INTELIGENCIA ARTIFICIAL

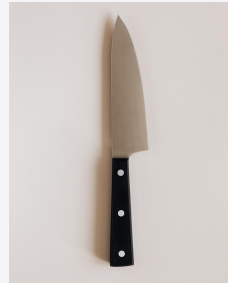
CONSIDERACIONES E IMPACTO SOCIAL

EN ESTA CLASE VAMOS A HABLAR DE

- Ética en la Inteligencia Artificial
- Sesgos y decisiones
- Buenas prácticas
- El lado bueno de la Inteligencia Artificial

UNA SERIE DE CONSIDERACIONES BÁSICAS SOBRE ÉTICA

- **Def.** La **ética** es la parte de la filosofía que trata del bien y del fundamento de sus valores
- La Inteligencia Artificial es una rama de conocimiento y per se el conocimiento nunca es malo, sin embargo es posible hacer un mal uso de él



EXISTEN MUCHOS PROYECTOS QUE BUSCAN ESTUDIAR LA ÉTICA EN TORNO A LA IA

- **Partnership on AI** es una organización sin fines de lucro que reúne a expertos (en IA y en derechos humanos) y empresas para establecer políticas de control y buenas prácticas
- **EthicNet** busca crear un conjunto de datos etiquetado por humanos que permita entrenar modelos que tomen consideraciones éticas
- **AI4People** es una organización europea que genera reportes e informes sobre el desarrollo de políticas en el ámbito de la IA
- **Algorithmic Justice League** se centra en reducir los sesgos de los modelos logrando modelos más éticos y justos

UNO DE LOS MAYORES PELIGROS EN LA IA ES LA APARICIÓN DE SESGOS

- **Def:** Un **sesgo** no es más que una predisposición hacia ciertas ideas o creencias que pueden reducir nuestra objetividad
- Todas las personas tenemos sesgos sociales que hemos ido absorbiendo por nuestra educación y nuestras experiencias
- Los sesgos suponen un problema cuando afectan a decisiones que deberían ser objetivas

EXISTEN DISTINTOS TIPOS DE SESGOS

- Hablamos de **sesgos en los datos** cuando empleamos un conjunto de datos que de manera consciente o inconsciente no transmite una visión amplia de la realidad. Por ejemplo, entrenamos un modelo de reconocimiento facial solo con fotos de una determinada raza

EXISTEN DISTINTOS TIPOS DE SESGOS

- Hablamos de **sesgos sociales** cuando nuestro modelo absorbe sesgos que se encuentran ya presentes en la sociedad. Por ejemplo, uno de los primeros modelos de lenguaje natural respondía a la analogía “hombre es a médico, lo que mujer es a ...” con “enfermera”



EL PROBLEMA DE ESTOS SESGOS SURGE AL APLICAR LOS MODELOS EN EL MUNDO REAL

En EEUU un banco liberó un modelo de sanción en el que había un sesgo racial derivado de manera implícita de la variable código postal. El modelo tendía a penalizar a personas residentes en barrios con predominancia de una determinada raza cuando realmente podían ser buenos pagadores



EN EUROPA CONTAMOS CON LEGISLACIÓN QUE NOS PROTEGE

- La legislación está aún en desarrollo pero las áreas principales como salud o finanzas ya han sido (al menos parcialmente) cubiertas
- El problema anterior nunca habría ocurrido en la UE porque el consumidor tiene, por ejemplo, derecho a que el banco le justifique de manera clara la denegación de acceso a la financiación

EN EUROPA CONTAMOS CON LEGISLACIÓN QUE NOS PROTEGE

- Herramientas más nuevas como ChatGPT se encuentra aún en un ligero limbo aunque los legisladores desarrollan poco a poco nuevas instrucciones y leyes

ALGUNAS BUENAS PRÁCTICAS PARA GARANTIZAR UN USO ÉTICO DE LA IA

1. Indicar aquellos mensajes o contenidos (imágenes, vídeos...) que se han generado mediante Inteligencia Artificial
2. No entrenar modelos con información sensible o datos no cedidos de **manera explícita**
3. Testar los modelos con datos que abarquen la variedad y diversidad de la población sobre la que se inferirá
4. Revisar de manera periódica los modelos y mantener canales de comunicación con los usuarios que los utilicen
5. Emplear en la medida de lo posible modelos explicables

LA INTELIGENCIA ARTIFICIAL TIENE IMPACTOS MUY POSITIVOS EN LA SOCIEDAD

- El uso de la IA y su democratización ha supuesto una revolución a nivel social
- La toma de decisiones basada en datos ayuda a reducir brechas generadas por la discriminación
- La implantación de la IA ha supuesto una mejora en la vida de todos

IA COMO APOYO CLAVE A LA ACCESIBILIDAD

- **Def.** La **accesibilidad** hace referencia a todas las técnicas para que personas con distintas dificultades puedan utilizar de manera más fácil distintas herramientas
- La IA ha supuesto una revolución este dominio con usos como:
 - Generación de descripciones de imágenes
 - Generación automática de subtítulos
 - Generación de texto y seguir instrucciones a partir de movimientos de ojos
 - Dispositivos de monitoreo de salud o valores concretos

IA COMO AYUDA A LA INVESTIGACIÓN Y AL CONOCIMIENTO

- La IA ha permitido generar y desarrollar nuevas ramas de investigación que han revolucionado prácticamente todos los campos científicos
- Se pueden destacar algunos ejemplos:
 - Asistencia de ChatGPT para revisiones bibliográficas
 - Avances en desciframiento de genomas mediante redes neuronales identificando patrones y mutaciones
 - Análisis de imágenes y patrones en temas como astronomía, geología, análisis de radiografías...
 - Abarca todos los campos desde las matemáticas a la salud o las Ciencias Sociales

OTROS CAMPOS EN LOS QUE LA IA HA TENIDO UN IMPACTO SIGNIFICATIVO

- **Salud y Medicina.** Permite una asistencia médica más personalizada y es una herramienta crucial para la detección temprana de enfermedades
- **Conducción autónoma.** Su implantación reducirá potencialmente los accidentes de tráfico y mejorará la eficiencia y accesibilidad
- **Mejoras en educación.** Acerca los recursos a los estudiantes y permite la construcción de planes de estudios más personalizados y adaptados
- **Asistencia en traducción.** Ha permitido la mejora de la comprensión y comunicación entre culturas así como democratizado el acceso a información/recursos/literatura

EN ESTA CLASE HEMOS HABLADO DE

- Implicaciones ética de la Inteligencia Artificial
- Fuentes de sesgo e importancia de su control
- Buenas prácticas cuando usamos Inteligencia Artificial / construimos modelos.
- El lado positivo y no tan conocido de la Inteligencia Artificial.