



Curso de Ética y Manejo de Datos para Data Science e Inteligencia Artificial

Sílvia Ariza Sentís



¿Para quién es este curso?



Personas expertas en análisis que quieren implementar un conocimiento y reflexión ética en sus análisis.



Personas que quieran aprender sobre IA y ciencia de datos teniendo conocimiento ético previo.



Entusiastas de los cambios sociales a partir de los datos.



**¿Qué son los datos
personales?**



Datos personales

Información que permite identificar a una persona de manera directa o indirecta.

Ejemplos:

- Edad
- Teléfono
- Domicilio
- Correo electrónico
- Documento nacional de identidad
- Número de seguridad social
- Ingresos
- Fecha de nacimiento





Clasificación de los datos personales

Por su jerarquía de riesgo/importancia:

- Datos personales ordinarios: nombre, dirección, finanzas personales.





Clasificación de los datos personales

- Datos personales sensibles:
 - Origen étnico
 - Opiniones políticas y religiosas
 - Orientación sexual
- Datos especiales:
 - Genéticos
 - Biométricos
 - Salud



**Datos que no se
consideran personales**





Datos NO personales

- Información anonimizada.
- Datos de personas jurídicas.
- Información que no permita la identificación.



**¿En qué se
diferencian de los
datos biométricos?**





Datos biométricos

Son un tipo de datos personales.
Permiten identificar digitalmente a una persona.

Ejemplos:

- Reconocimiento facial
- Reconocimiento de retina
- Reconocimiento vascular
- Reconocimiento de firma
- Reconocimiento de escritura
- Reconocimiento de voz
- Reconocimiento de escritura de teclado





Clasificación de los datos biométricos

Por naturaleza:

- Universal
- Único
- Permanente



Según sus características:

- Rasgos físicos y fisiológicos.
- Rasgos de comportamiento y personalidad.

¿En qué pensamos cuando hablamos de regulaciones y ética de la información?





Privacidad

Blockchain

Anonimizar

Datos personales

Protección

GDPR

Leyes

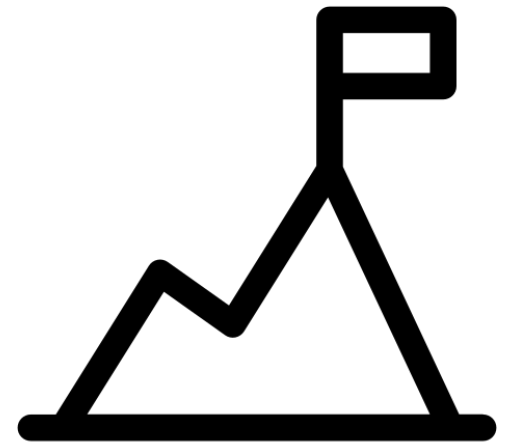
Cookies



Reto

¿Qué otros datos
personales conoces?

Encuentra su
tipología.





Escándalos históricos



Importancia de la privacidad

1940, Holanda

El **1.5%** de la población de Holanda en 1940 era judía.

La mayoría eran familias establecidas en el país por siglos.

Los nazis se apoderaron del censo del gobierno holandés y pudieron acceder a toda la información.

75% de los judíos holandeses fueron asesinados por las fuerzas nazis.





Importancia de la privacidad 2016, México

- Los datos personales no están bien protegidos.
- Los datos del INE aparecieron en un repositorio S3 abierto de Amazon (también en Ecuador e India).



Otros casos:

Usuarios de **Yahoo** (3 billones).

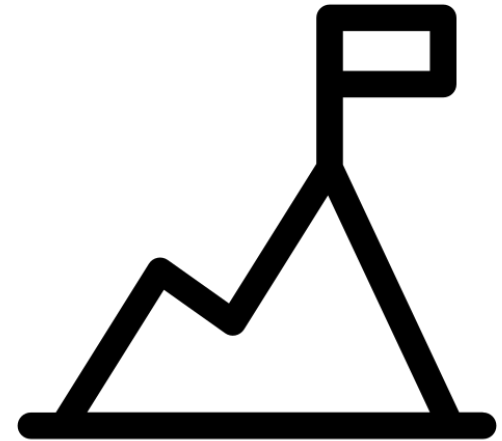
Clientes de **Marriott** (500 millones).

Cuentas de **Facebook** (533 millones).



Reto

¿Qué otro escándalo histórico o actual relativo a la falta de privacidad y/o fuga de información conoces?





**Interés creciente
por la información**



Las empresas tienen más interés por los datos

Recolectar información es cada vez más **económico**.

Las empresas entienden el **valor** de entender a su usuario.





Las empresas tienen más interés por los datos

Hay un alto **incentivo** para almacenar información, pero tiene que **regularse** su uso.

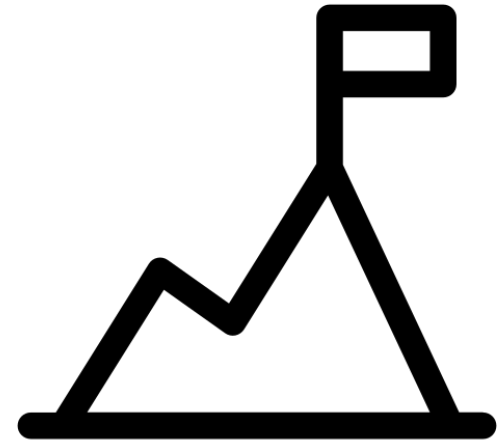
Algunas empresas se dedican a la **compra** y **venta** de bases de datos.





Reto

¿Qué otras razones
pueden motivar una
institución a recolectar
información?





Bias y GIGO en datos



Bias (sesgo)

Sucede cuando seleccionamos la información de manera errónea y considerando **una muestra que no representa la totalidad de la población**. En este caso, la muestra aleatoria no es representativa.

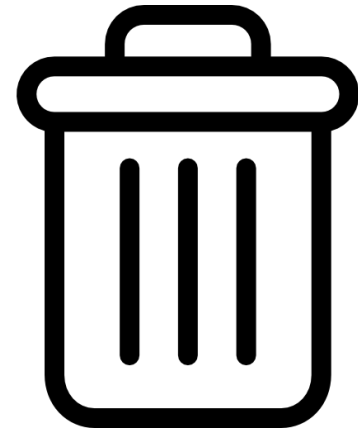
Ejemplo: sesgo racial.





Garbage in, garbage out

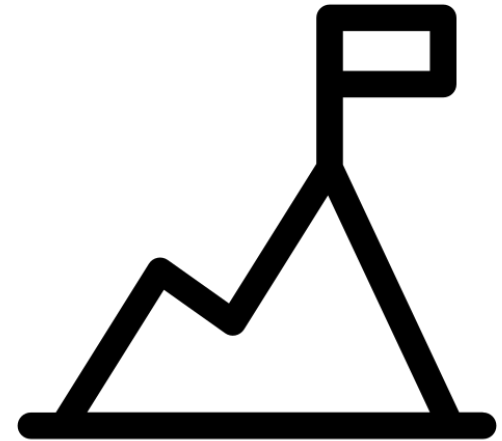
La calidad del
resultado (output)
depende de la calidad
de la entrada (input).





Reto

¿Qué otros tipos de sesgos crees que suceden en las empresas?





Advertencias de uso en marketing

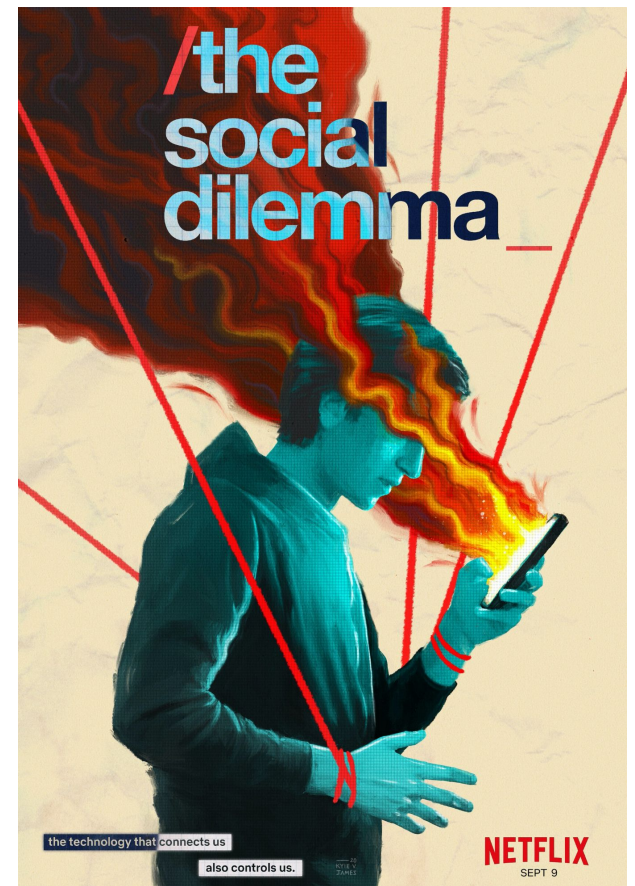


Alertas de uso de la información

Podemos crear publicidad personalizada.

Advertencias:

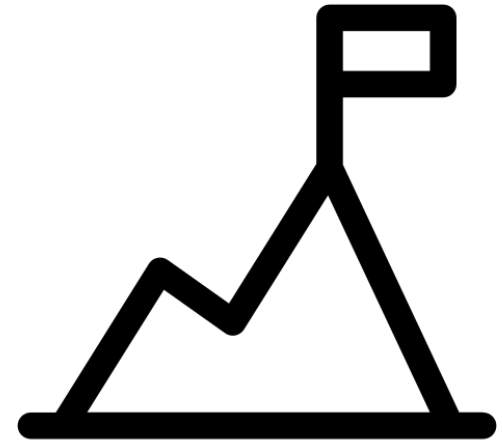
- a) Sesgos **discriminatorios**.
- b) **Precios** por información privilegiada.
- c) **Manipulación** de campañas políticas.
- d) Fomento de conductas **adictivas** y ataques **emocionales**.





Reto

Piensa en empresas u organizaciones que históricamente han infringido alguno de esos puntos.





Advertencias de uso en campañas políticas



El escándalo de Cambridge Analytica

Cambridge Analytica ofrecía a negocios y partidos políticos la posibilidad de *cambiar/manipular el comportamiento de la audiencia*.

- a) Datos propios a partir de encuestas.
- b) **Datos de redes como Facebook.**
- c) **Ciencia del comportamiento y modelos predictivos.**





El escándalo de Cambridge Analytica

Se identificaron personas que podían ser targets de campañas de marketing para incidir en sus preferencias políticas.

Campañas:

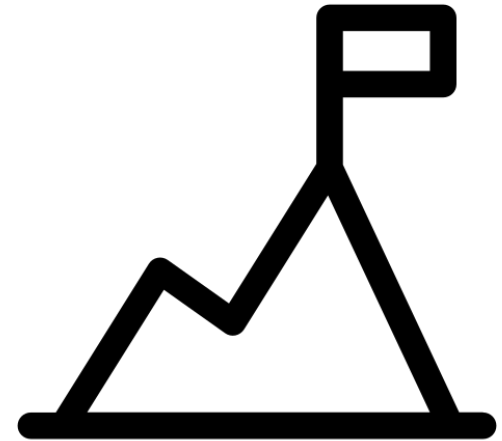
- Donald Trump en EEUU.
- PRI en México.





Reto

Reflexiona y
escribe una
potencial solución
a este problema.





Reglamento



Ley Federal de Protección de Datos Personales en Posesión de Particulares

- En México se aplica la ley contra las empresas que no protejan la información de sus clientes de manera adecuada.
- Dato personal: información que permite identificar a una persona.



Ley Federal de Protección de Datos Personales en Posesión de los Particulares y su Reglamento



Ley Federal de Protección de Datos Personales en Posesión de Particulares

- Obliga a las empresas a informar del uso que le darán a la información recolectada.
- Sanciona la compra de bases de datos personales.



Ley Federal de Protección de Datos Personales en Posesión de los Particulares y su Reglamento



Reglamento General de Protección de Datos (GDPR)

La ley de la Unión Europea regula la protección de datos y su privacidad para todos los ciudadanos de la UE y de la AEE así como la transferencia de datos fuera de la región.





Reglamento General de Protección de Datos (GDPR)

Pretende dar control a los ciudadanos sobre sus propios datos y simplificar el entorno regulatorio.

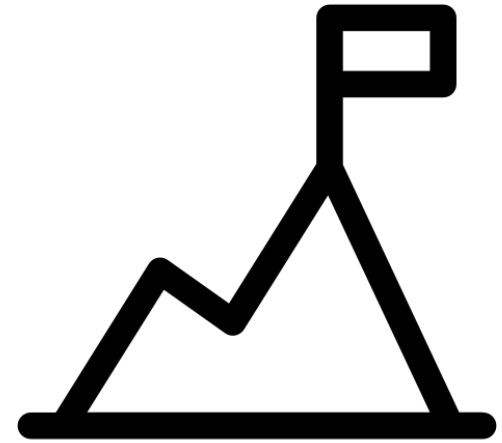
La ley aplica para cualquier empresa del mundo que procese datos de personas residentes en la UE.





Reto

¿Cuál es tu opinión sobre la implementación de reglamentos en datos?





Ética y deep learning: vehículos autónomos



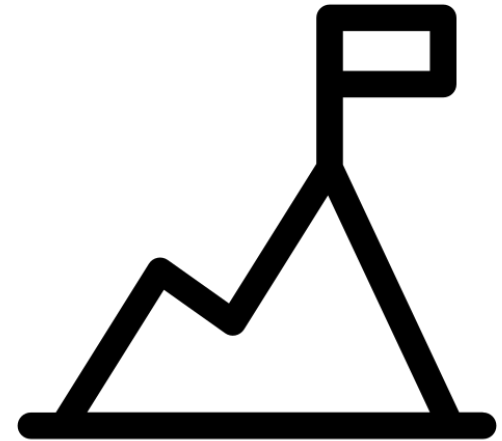
Vehículos autónomos





Reto

¿Crees que ya deberíamos disponer de esta tecnología en la calle?





Ética y deep learning: reconocimiento facial



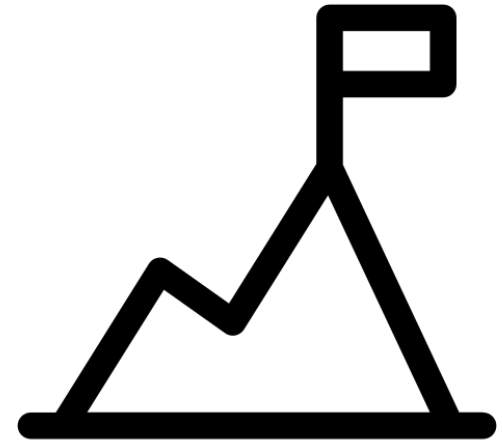
Reconocimiento facial





Reto

¿Qué opinas sobre esta tecnología, en qué casos crees que es seguro o cómodo su uso?





Ética en la pandemia



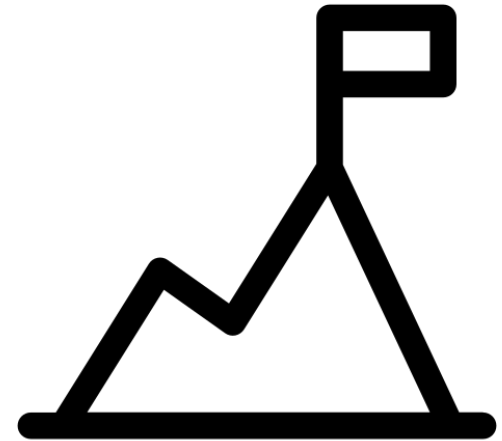
Pandemia





Reto

¿Crees que este uso de los datos facilita la información o el control?





Ética en las relaciones interpersonales



Relaciones interpersonales





Funciones de audio





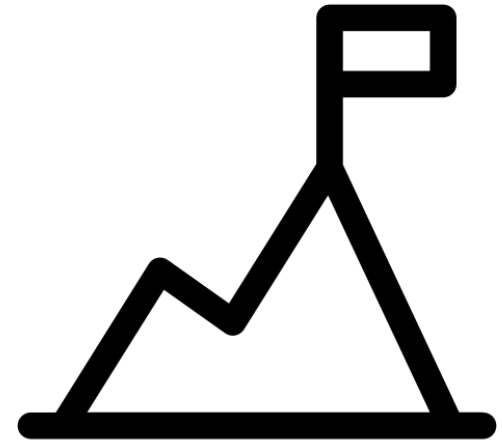
Redes sociales después de la vida





Reto

¿Cómo te sientes más cómodo o cómoda? ¿Cuál crees que es la línea entre la interacción virtual y presencial?





Ética y Procesamiento del Lenguaje Natural (NLP)



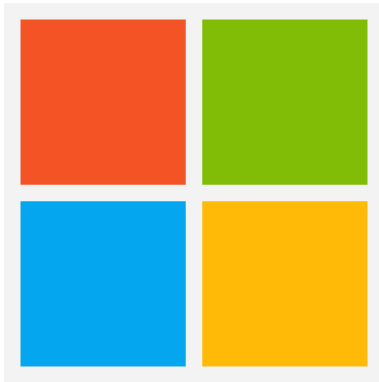
Procesamiento del Lenguaje Natural (NLP)

Investigación de la interacción entre las computadoras y el lenguaje humano mediante el uso de **lenguas naturales**.





Procesamiento del Lenguaje Natural (NLP)



vs.





La inteligencia artificial supera al ser humano



VS.



La IA de Microsoft y Alibaba supera a estudiantes de Stanford en test de comprensión lectora.

Reto: identidad / emoción.



GPT-3

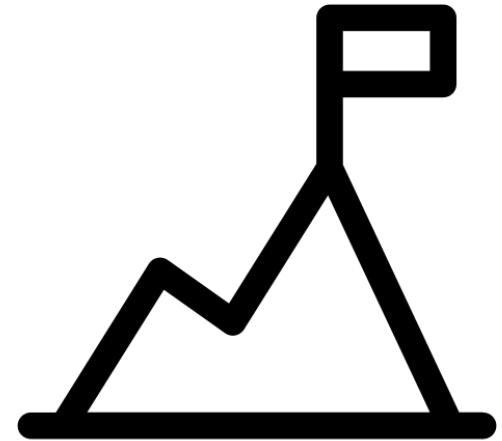
Lenguaje autorregresivo que emplea el aprendizaje profundo (deep learning) para producir textos que simulan la **redacción** humana.





Reto

¿Crees que el NLP pueda alcanzar el conocimiento y sensibilidad humanos?





**¿Qué son las
políticas públicas?**



Políticas públicas

- Atención del Estado a una necesidad social.
- Implementación a partir de normas, instituciones, prestaciones y servicios.
- Ayudan a alcanzar principios éticos e igualdad de derechos.



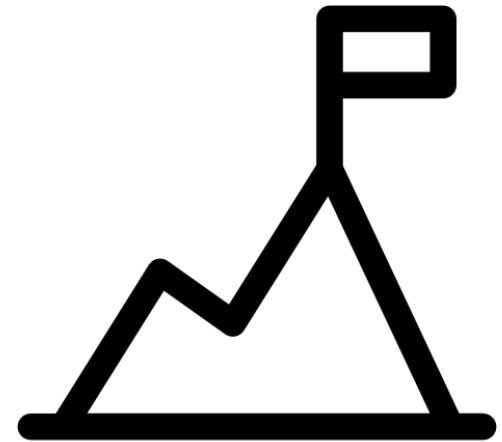
Políticas públicas: tipología

- **Autoridad institucional** (municipal, nacional,...).
- **Sector** (educación, medio ambiente,...).
- **Destinatarios** (personas con discapacidades, jóvenes,...).
- **Elaboración** (participativa o autoritaria).
- **Planificación** (anticipativa o reactiva).



Reto

Piensa en políticas públicas que se hayan desarrollado en tu ciudad o país.





Datos y prevención de crímenes



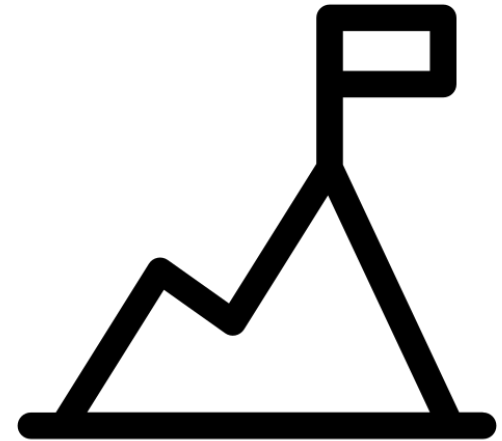
Prevención de crímenes





Reto

¿Qué otras aplicaciones de prevención de crimen podrían lograrse con datos e inteligencia artificial?





Datos y salud



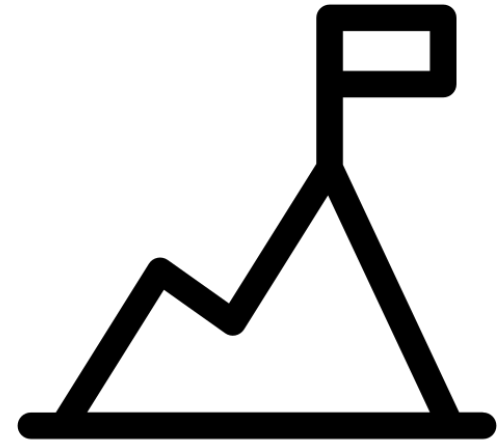
Salud





Reto

¿Qué otras aplicaciones de prevención de enfermedades y accidentes podrían lograrse con datos e inteligencia artificial?





Datos y movilidad



Movilidad





Ciudades inteligentes (Smart Cities)





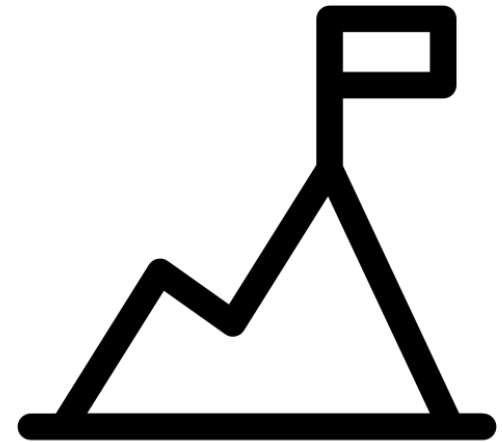
La otra cara de la moneda





Reto

¿Qué opinas del potencial sesgo de rutas y discriminación de las plataformas de movilidad?





Datos y educación



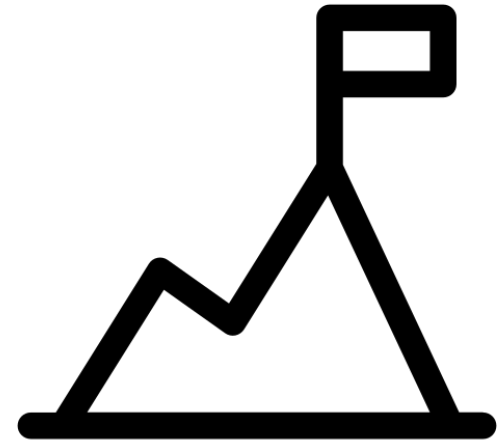
Educación





Reto

¿Qué otras aplicaciones de mejora educativa podrían lograrse con datos e inteligencia artificial?





Datos y medio ambiente



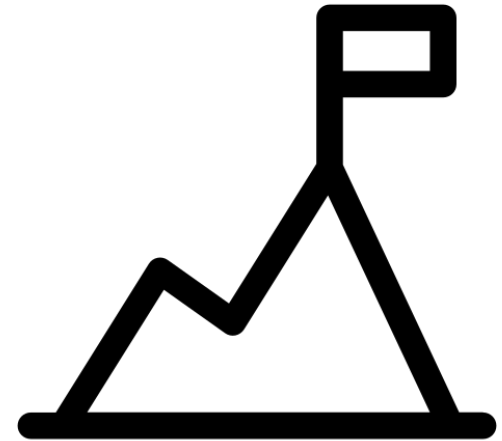
Medio ambiente





Reto

¿Qué otras aplicaciones de mejora al medio ambiente podrían lograrse con datos e inteligencia artificial?





Conclusiones y reflexiones



¡Muchas felicidades!

- Completar tu guía de reflexiones y retos.
- Aprobar el examen.
- Compartir qué te pareció el curso en tu reseña.

