

Sesión

5

Aprendizajes esperados

Duración sugerida

Al final de esta sesión se espera que puedas:



Explicar la importancia de colaborar con personas diferentes a sí mismo (sexo, edad, raza, cultura, etc.) en proyectos computacionales.



40%

40%

20%



Explicar qué es el sesgo algorítmico.



Identificar en los datos de entrenamiento una de las causas del sesgo algorítmico.

Material para la clase

- Anexos 5.1, 5.2 y 5.3.
- Computador con acceso a internet.
- Anexos 5.4, 5.5, 5.6 y 5.7.

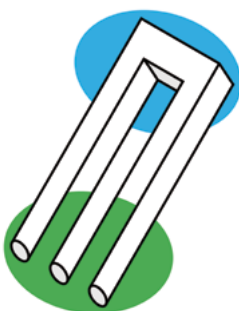


Anexos

Anexo 5.1



Anexo 5.2



Lo que sabemos, lo que debemos saber



Esta sección corresponde al 40% de avance de la sesión



¿Recuerdas el ejemplo que se mostró en una sesión pasada en la que un modelo entrenado para identificar una mano levantada no pudo identificarla?

Quizás te haya sucedido algo similar con alguno de tus modelos. En el caso del ejemplo, los datos de entrenamiento solo tenían imágenes con la mano izquierda levantada. Cuando se le presentó al modelo un dato de prueba con la mano derecha levantada no logró clasificarla como mano levantada, a pesar de que sí había una mano levantada. En esta sesión vamos a profundizar en este tipo de situaciones, entendiendo cómo se presentan y cómo se pueden prevenir y corregir.

Para empezar, observa en silencio las imágenes que te muestra tu docente y responde las preguntas que irá haciendo en caso de que te dé la palabra.

Primero, tu docente mostrará la imagen del Anexo 5.1 y hará las siguientes preguntas a un par de estudiantes para escuchar sus diferentes puntos de vista:



¿Qué ven en la imagen?

¿Pueden ver ustedes lo mismo que estos estudiantes ven?

Luego, tu docente mostrará la imagen del Anexo 5.2 y repetirá las mismas preguntas.

Anexo**Anexo 5.3**

Finalmente, tu docente mostrará la imagen del Anexo 5.3 y hará, con respecto a esa última imagen, las siguientes preguntas de la misma manera que hizo las anteriores:



¿De qué color es el vestido?

¿Pueden ver lo que sus compañeras y compañeros de clase ven?

Y con respecto a todas las imágenes:



Si alguna persona vio en alguna de las imágenes algo diferente de lo que tú viste, ¿quiere decir que alguno de los dos está equivocado?

¿Habrá una única respuesta correcta?

¿Cómo te sientes cuando alguien no está de acuerdo con lo que tú ves?

Lo que percibimos, incluyendo lo que vemos, oímos, sentimos está basado en nuestra **perspectiva**, es decir, en nuestro punto de vista.

Ilustremos un ejemplo. Como quizás sabes un perro tiene mejor olfato que los seres humanos, por lo que un olor característico que el perro identifique puede ser algo inoloro para ti. Esto genera perspectivas diferentes sobre el mismo hecho, sin que ninguno de los dos esté bien o mal. A lo largo de la historia las personas hemos aprovechado esta diferencia de perspectiva con los perros, por ejemplo, al entrenarlos para cazar.



¿Puedes pensar por qué es importante entender las perspectivas de las otras personas?

¿Tendrá solo que ver con cómo se sienten (recuerda tu respuesta a la pregunta 3 anterior) o habrá alguna ventaja en hacerlo?



Dada la importancia de tener en cuenta y de entender la perspectiva de otras personas, ¿qué podrías hacer para lograrlo?

¿Será que tú sola (solo) puedes fácilmente darte cuenta y entender otras perspectivas?

Ahora, volviendo a la IA, recuerda los modelos que entrenaste para clasificar posturas y los números uno, dos y tres. Trae a tu memoria también el ejemplo de posturas presentado en la guía en el que la IA no podía identificar la mano levantada si no era la misma con la que había sido entrenada (suponiendo que solo fue entrenada levantando una mano).

Responde:



¿Crees que las máquinas con IA tienen una perspectiva?

Como viste con las imágenes que te mostró tu docente, si nos quedamos con una sola perspectiva puede que no estemos considerando todos los aspectos y perspectivas. Cuando esto sucede, cuando nos aferramos a una sola perspectiva es más probable que cometamos errores como el **sesgo**.



¿Se te ocurre algún ejemplo de sesgo?

Los sesgos son muy comunes y son parte de nuestra propia naturaleza. A menudo nos sirven o son inocuos, pero otras veces nos afectan. Te damos un ejemplo:



Si en nuestros contextos se dice que cierta fruta es muy buena para la salud, dado que cura todo, es altamente probable que cuando se nos pregunte sobre el tema respaldemos esta idea o que cuando alguien nos presente un estudio que dice que ello es mentiras, simplemente consideremos que está equivocado ese estudio.

Estos sesgos suceden también con las personas cuando hemos escuchado cosas malas de un cierto grupo humano. Tenderemos a tomar decisiones considerando que ese grupo humano es malo, sin ninguna razón o evidencia.



¿Cómo te sentirías si alguien está sesgado, es decir, tiene un sesgo en tu contra?

Como ya has visto, la IA aprende a partir de datos de entrenamiento, ¿qué pasa si esos datos contienen sesgos? ¿La IA tomará decisiones sesgadas?

En la siguiente actividad vamos a ver si es así.

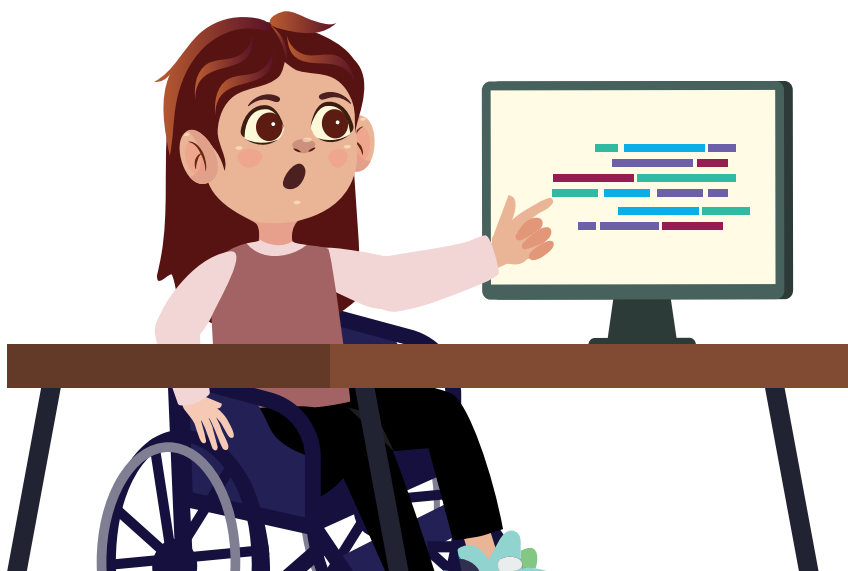
Glosario



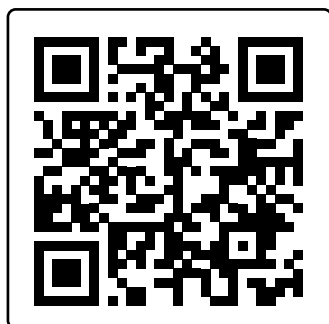
Perspectiva: enfoque o visión desde la cual se observa y analiza un problema, influenciando las decisiones sobre el diseño del modelo, la selección de datos y las técnicas de entrenamiento.



Sesgo: es un atajo mental que nos lleva a aceptar algo o tomar una decisión de forma rápida sin necesidad de pensar mucho. En muchos casos de la vida cotidiana nos son útiles, pero frente a grandes decisiones en situaciones complejas nos lleva a tomar malas decisiones que nos afectan o afectan a otras personas.

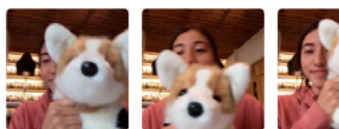


Enlace



Teachable Machine

Figura 1. Proyecto de imagen



Proyecto de imagen

Realiza la preparación con imágenes de archivos o de la webcam.

Figura 2. Modelo de imagen estándar

Nuevo proyecto de imagen

Modelo de imagen estándar

Ideal para la mayoría de usos

Imágenes a color de 224 x 224 px

Exportar a TensorFlow, TFLite y TF.js

Tamaño del modelo: alrededor de 5 MB

Manos a la obra

Conectadas



Esta sección corresponde al 80% de avance de la sesión

Vas a volver a entrenar un modelo de IA creado en *Teachable Machine*, pero esta vez para clasificar imágenes de perros y gatos. Como ya has usado esta herramienta de IA en dos ocasiones, en esta actividad se irá un poco más rápido. Si quieres ir paso por paso, puedes regresar a la guía de la sesión anterior.

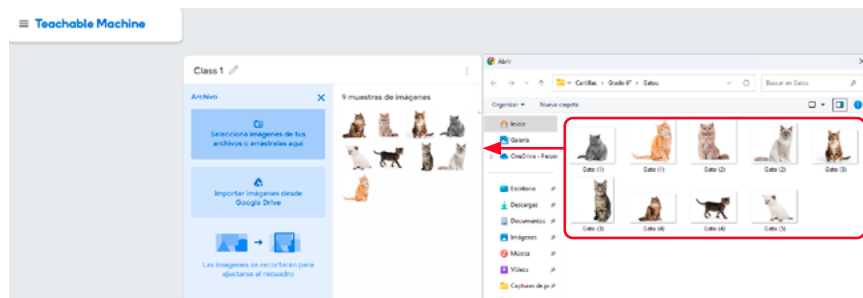
Organízate en grupos de 2 a 3 personas por computador. Lo primero es descargar y descomprimir el conjunto de datos que se va a usar, que se encuentra en el enlace al principio de la guía. En caso de necesitar ayuda, pide asistencia a tu docente.

Ahora ve a *Teachable Machine* escaneando o haciendo clic sobre el código QR al inicio de esta sección.

En las sesiones anteriores se trabajaron el proyecto de postura y el de audio. Ahora van a hacer el **Proyecto de imagen**, ver *Figura 1*. Después de haber hecho clic en este, seleccionen **Modelo de imagen estándar**, ver *Figura 2*.

Cambien el nombre de la primera clase a Gato. En esta ocasión las imágenes no se tomarán con la cámara web, sino que son archivos que van a subir, así que seleccionen **Subir** en la clase Gato. Luego, abran la carpeta que descargaron y descomprimieron, *Teachable Machines Cat-Dog Dataset*, y abran la carpeta *cat training dataset* que se encuentra en su interior. Seleccionen todas las imágenes y arrástralas al recuadro que dice algo como **Elija imágenes de sus archivos** o arrástralas y suéltelas aquí, como se muestra en la *Figura 3*.

Figura 3. Elegir y subir imágenes al modelo. Clase Gato



Anexo

Anexo 5.4

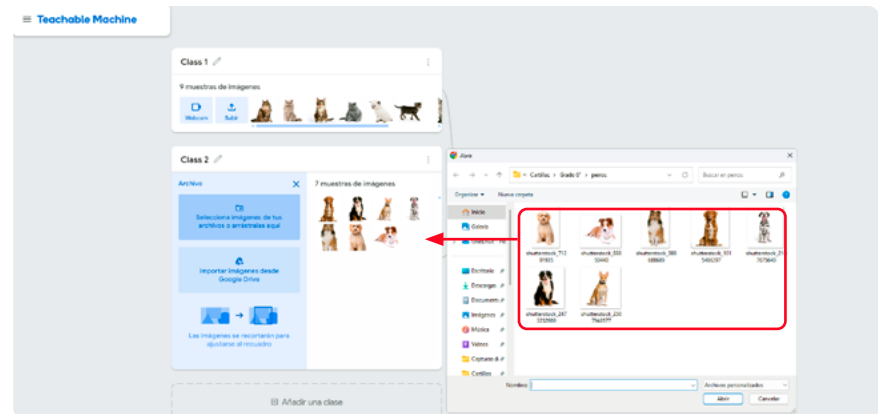
Datos de entrenamiento	Número de imágenes	Número de posturas diferentes ¹	Número de características diferentes ²
Gatos			
Perros			
Datos de prueba	Clasificado como (gato o perro)	Nivel de certeza del modelo (%)	Acierto al modelo (sí o no)
			
			
			
			
			

1. ¿Cuáles imágenes consideran que son bien diferentes con respecto a la postura de la mascota? Por ejemplo, si hay 2 imágenes con gato que están sentados y otras 3 con gato acostados para el mismo foto, podrían decir que en esas 5 imágenes hay solo 2 imágenes diferentes, con respecto a la postura. Aquí no hay respuestas correctas e incorrectas, es más una opinión. Discute y argumenta con tu grupo para llegar a un acuerdo.

2. Es igual que el número de posturas, pero con respecto a las características de la mascota, como color o colores, longitud del pelo, contornos, etc. Por ejemplo, si hay 4 imágenes con perro que al mirarlo rápidamente se ven como de colores diferentes: contornos sencillos o longitud del pelo sencillos consideras como 1 solo.

Ahora repitan el mismo proceso con la segunda clase que tiene los perros. En este caso la subcarpeta se llama *dog training data*. Se debe ver similar a la que aparece en la Figura 4.

Figura 4. Elegir y subir imágenes al modelo. Clase Perro



Como grupo, llenen la fila correspondiente a los datos de entrenamiento de la clase Perro del Anexo 5.4.

Si ya terminaron de subir el conjunto de datos de entrenamiento, es momento de entrenar su modelo. Una vez entrenado, se debe ver algo como lo que se muestra en la Figura 5.

Figura 5. Ensayo modelo proyecto de imagen

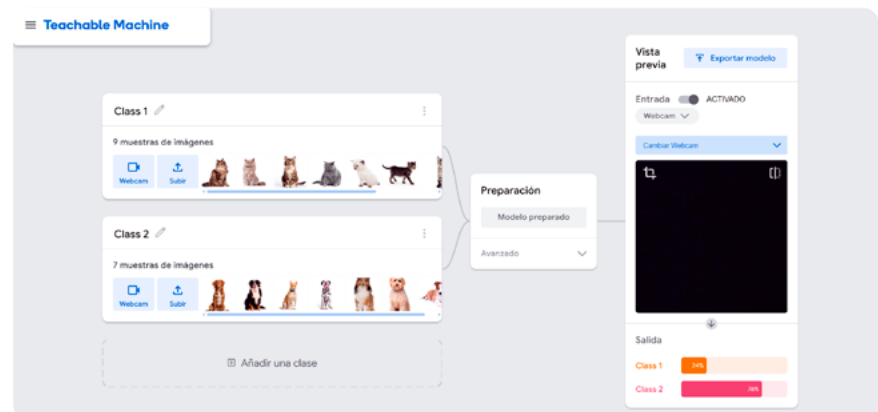
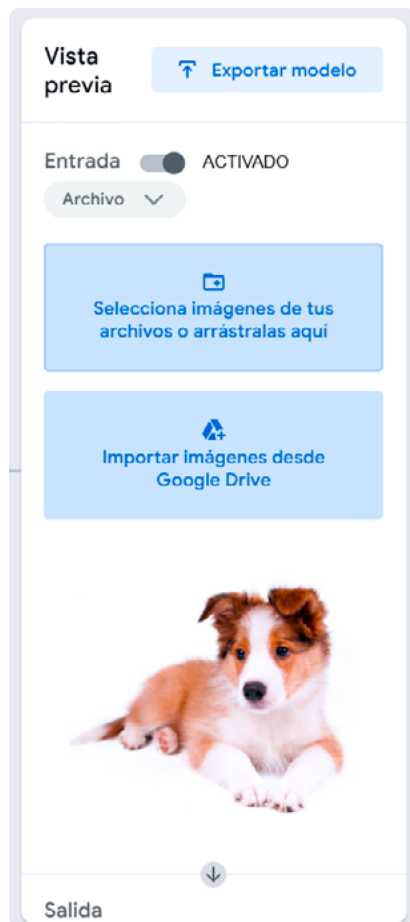


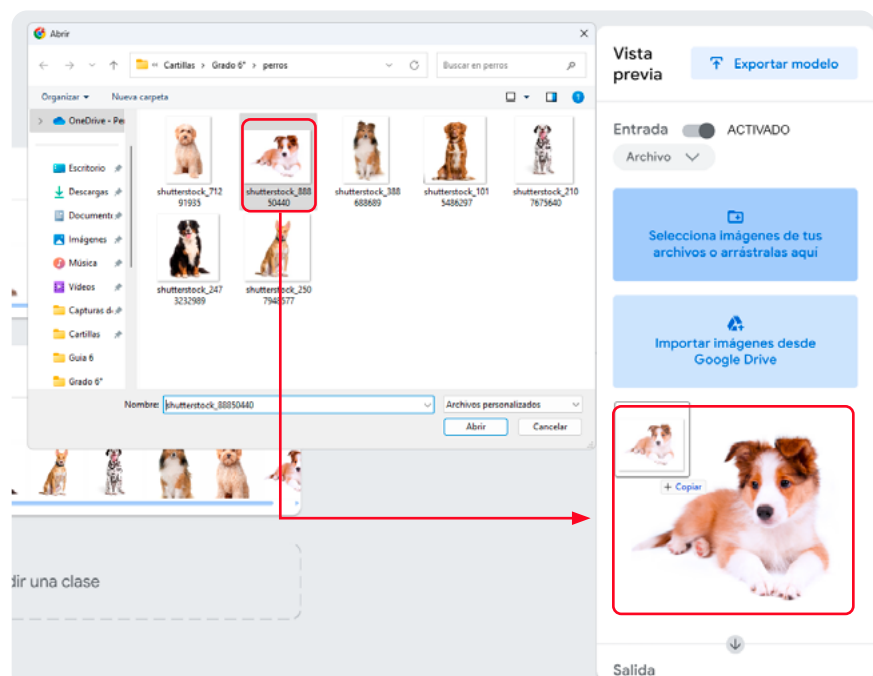
Figura 7. Ensayo modelo proyecto de imagen.



Los datos de prueba, al igual que los datos de entrenamiento, son imágenes que se encuentran en la subcarpeta *test dataset* que descargaron antes.

Para que el modelo reciba archivos como datos de prueba deben hacer clic donde dice **Cámara Web** y seleccionar **Archivo**. Luego deben arrastrar una imagen de los datos de prueba al recuadro que dice algo como **Selecciona imágenes de tus archivos** o arrástralas aquí. Debería verse como aparece en la *Figura 6*.

Figura 6. Cargue de datos de prueba



Una vez suelten la imagen, el modelo debe mostrarles la imagen que está evaluando y su clasificación en la parte inferior, como se ve en la *Figura 7*. Al igual que en la sesión anterior aparecen las barras, cuya longitud representa la certeza del modelo en la clasificación de cada una de las clases. Sin embargo, en este caso, como son imágenes estáticas, las barras se quedan quietas y es posible ver también un porcentaje que corresponde a ese **nivel de certeza**, también conocido como **grado de confianza**, del modelo.

Anexos

Anexo 5.5

Nombre: _____

1. ¿Les sorprendió alguno de los resultados? ¿Cuáles y por qué?

2. ¿Se equivocó el modelo clasificando alguno de los datos de prueba? ¿En cuáles?

3. ¿Por qué creen que se equivocó?

4. ¿En qué clase funcionó mejor el modelo?

5. ¿Por qué creen que funcionó mejor en esa clase?

6. ¿Consideran que el modelo tiene un sesgo algorítmico? ¿A qué clase privilegia?

7. ¿Qué podrían hacer con el conjunto de datos de entrenamiento para mejorar el modelo?

Anexo 5.6

Datos de entrenamiento	Número de imágenes	Número de posturas diferentes ¹	Número de características diferentes ²
Gato			
Perrito			
Datos de prueba	Clasificado como (gato o perrito)	Nivel de certeza del modelo (%)	Acierto al modelo (si o no)
			
			
			
			
			

1. ¿Cuáles imágenes consideran que son bien diferentes con respecto a la postura de la mascota? Por ejemplo, si hay 2 imágenes con gatos que están sentados y otras 3 con gatos acostados para el mismo lado, podrían decir que en esas 5 imágenes hay solo 2 imágenes diferentes, con respecto a la postura. Aquí no hay imágenes sentadas o acostadas, así que son 5 imágenes diferentes. Discute y argumenta con tu grupo para llegar a un acuerdo.

2. Es igual que el número de posturas, pero con respecto a las características de la mascota, como color o rasgos, longitud del pelo, contornos, etc. Por ejemplo, si hay 4 imágenes con perros que al mirarlos rápidamente se parecen, son de colores similares, contornos similares y longitud del pelo similar, podrían contar como 1 solo.



A medida que vayan realizando las pruebas, vayan llenando la tabla correspondiente a los datos de prueba del Anexo 5.4.

Cuando hayan terminado y usando lo que consignaron en el Anexo 5.4, discutan y respondan las preguntas 1 a 5 del Anexo 5.5.

Cuando un algoritmo, en especial de IA, tiene salidas (predicciones o decisiones) que no responden de forma adecuada a un contexto y la respuesta se da repetidamente, decimos que tiene un **sesgo algorítmico**. Al decir “injustas” se entiende que privilegian o discriminan a una predicción o decisión en particular. Es decir, no son errores al azar, sino que muestran una tendencia.

Una causa frecuente de sesgo algorítmico es un conjunto de datos de entrenamiento sesgado. Este puede no tener variedad y/o cantidad de datos similares en las diferentes clases.



Ahora, usando lo que escribieron en el Anexo 5.4, discutan y respondan las preguntas 6 y 7 del Anexo 5.5.

Es momento de corregir el sesgo algorítmico de su modelo. Pueden agregar nuevas imágenes a cada una de las clases de la misma forma como lo hicieron antes. Pueden usar las imágenes que encuentran en la subcarpeta extra cat-dog images.



Mientras lo hacen, llenen los datos de entrenamiento del Anexo 5.6.

Una vez hayan agregado las imágenes que consideren a la(s) clase(s) que consideren, deben volver a entrenar su modelo.

Anexo

Anexo 5.7

Nombre: _____

1. ¿Se equivocó el modelo clasificando alguno de los datos de prueba? ¿En cuáles? ¿Son los mismos de antes del ajuste de los datos de entrenamiento?

2. En términos generales, ¿mejoraron los niveles de certeza de las predicciones?

3. En conclusión, ¿fue lo que mejoró el modelo con el ajuste que hicieron? ¿Por qué?

4. ¿Todos los miembros del grupo tuvieron las mismas ideas durante esta actividad y opinaron lo mismo durante esta actividad?



A medida que van realizando las pruebas, vayan llenando la tabla correspondiente a los datos de prueba del Anexo 5.6.

Finalmente, respondan las preguntas del Anexo 5.7.

Glosario



Nivel de certeza: se expresa generalmente como un valor numérico, que indica qué tan seguro está el modelo sobre una clase o resultado determinado.



Grado de confianza: es una medida que indica la probabilidad de que una predicción realizada por el modelo sea correcta. Se expresa generalmente como un valor entre 0 y 1, donde un valor cercano a 1 indica alta confianza en la predicción.



Sesgo algorítmico: es la tendencia de un modelo a producir resultados sistemáticamente erróneos debido a prejuicios en los datos de entrenamiento, el diseño del algoritmo o las decisiones del desarrollador.



Antes de irnos



Esta sección corresponde al 100% de avance de la sesión

Revisa los aprendizajes esperados de forma individual, respondiendo las preguntas de forma que mejor reflejen tu progreso:

- 1 ¿Puedes explicar qué es el sesgo algorítmico?
 - ☐ Sí
 - ☐ Parcialmente
 - ☐ Aún no
- 2 ¿Puedes identificar en los datos de entrenamiento una de las causas del sesgo algorítmico?
 - ☐ Sí
 - ☐ Parcialmente
 - ☐ Aún no

Si tus respuestas fueron “Parcialmente” o “Aún no”, vuelve a las actividades propuestas en *Teachable Machine*. Luego discute con tus compañeras y compañeros de grupo lo que pasó al cargar los datos de entrenamiento y al hacer uso de los datos de prueba. Si todavía te quedan dudas, consúltale a tu docente.

Haz un gráfico que incluya lo que viste en esta sesión. Podría incluir:

Definiciones:

- ☐ Perspectiva
- ☐ Sesgo
- ☐ Sesgo algorítmico
- ☐ Nivel de certeza / grado de confianza

Conceptos:

- ☐ ¿Por qué es importante hacer una buena selección de los datos de entrenamiento de un modelo de IA?
- ☐ ¿Qué estrategias se pueden usar para hacer una buena selección de datos?
- ☐ ¿Puede estar sesgada la IA?
- ☐ ¿Cuál es una de las posibles causas de los sesgos de la IA?

Si necesitas volver a leer para hacer tu gráfico, no dudes en hacerlo.



La computación y la sociedad

En esta guía pudiste trabajar con herramientas de inteligencia artificial.

La inteligencia artificial tiene el potencial de ayudar a las personas y a la sociedad en muchas actividades; quienes las sepan utilizar tendrán oportunidades interesantes de colaborar en la solución de problemas complejos. Esto permite desarrollar soluciones de problemas complejos en áreas como la salud, la educación y el medio ambiente. Además, permite mejorar la eficiencia en tareas cotidianas, desarrollar innovaciones tecnológicas y crear soluciones que beneficien a la sociedad.