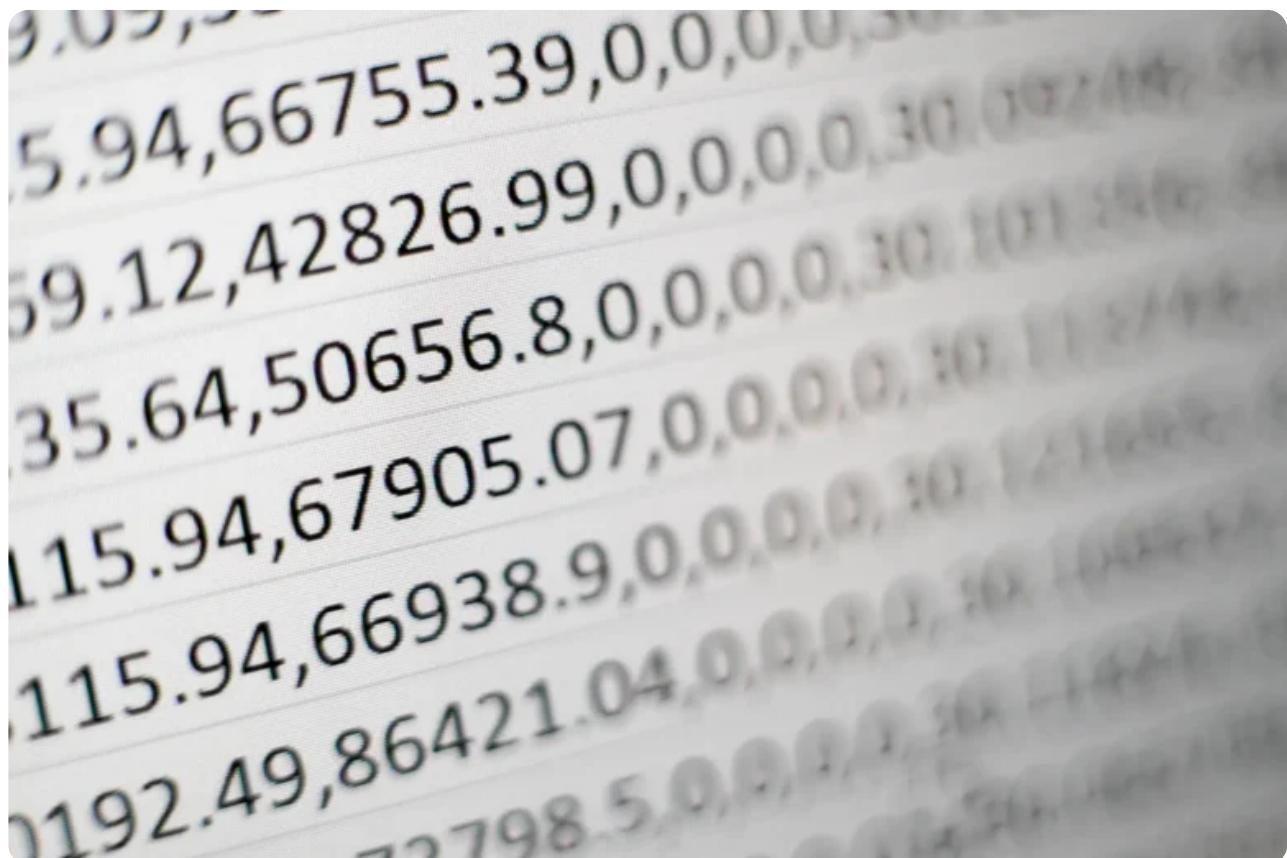


[Volver al blog de tareas](#)

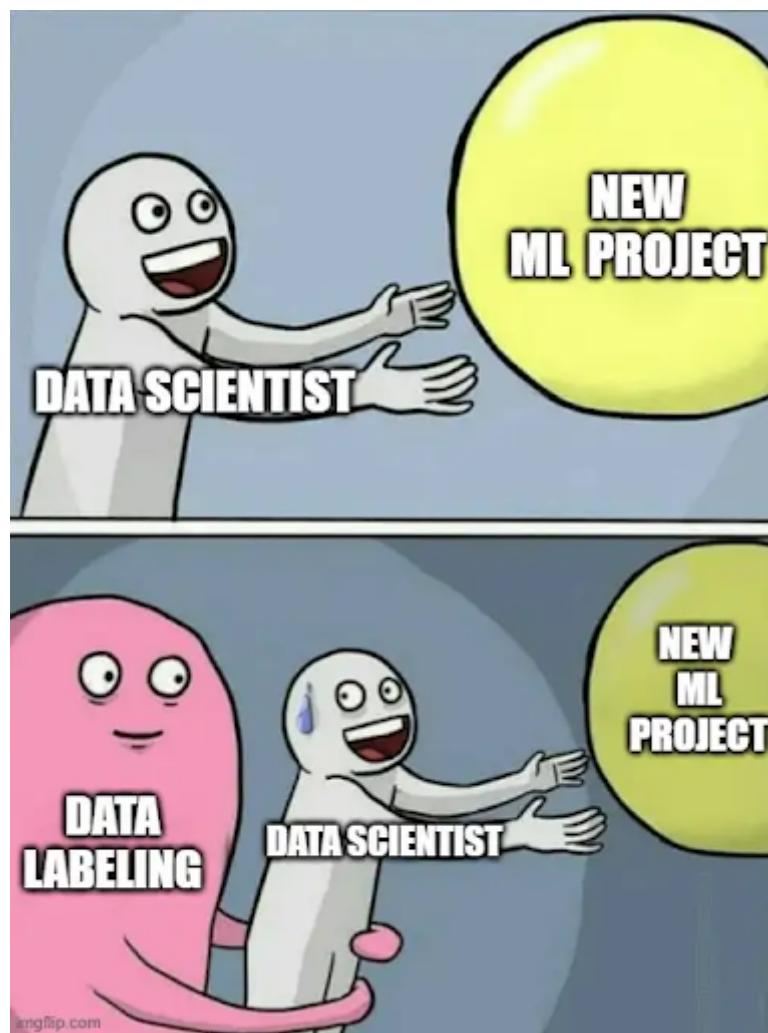
Etiquetado de datos automatizado frente a etiquetado de datos manual: ¿cuál es la diferencia?

Por: Nati Catalán

14 de diciembre | 2022



"La base para las tareas de aprendizaje automático supervisado son los conjuntos de datos etiquetados con precisión".



Los sistemas de aprendizaje automático dependen en gran medida de los datos. Estos sistemas son capaces de encontrar patrones ocultos y proporcionar la información necesaria relacionada con una determinada tarea. Pero los datos deben estar bien seleccionados, que es donde entran en escena tareas como **la estructuración y el etiquetado de los datos**.

Pero, ¿cómo podemos crear los conjuntos de datos precisos y confiables que exigen las tareas? En general, hay dos enfoques mediante los cuales respondemos esta pregunta: **manual y. etiquetado de datos automatizado o habilitado para IA**

Este artículo explora estos dos enfoques, es decir, el etiquetado manual frente al automático.

y te presenta los pros y los contras de cada uno de ellos. Además, también veremos brevemente un estudio de caso.

¿Qué son los datos etiquetados?

La tarea de etiquetado de datos para aprendizaje automático se utiliza en la tarea de aprendizaje supervisado. Para que el modelo aprenda de manera efectiva, las etiquetas deben ser precisas y completas. Debe reflejar con precisión las relaciones entre las características y las variables objetivo. Esto permite que el modelo de ML aprenda los patrones y las relaciones subyacentes en los datos y generalice esos patrones a datos nuevos e invisibles.

Además de proporcionar la [verdad básica](#) para los algoritmos de aprendizaje automático, el etiquetado de datos también se puede utilizar para mejorar el rendimiento y la confiabilidad de los modelos. Al seleccionar y etiquetar cuidadosamente los datos de entrenamiento, es posible garantizar que los modelos se entrenen en una muestra representativa y diversa de los datos, lo que puede ayudar a mejorar su precisión y capacidad de generalización.

Esencialmente, [el etiquetado de datos](#) es una parte integral del proceso de aprendizaje automático y es crucial para garantizar la efectividad y confiabilidad de los modelos de aprendizaje automático.

¿Cuáles son las diferentes tareas en el etiquetado de datos?

El etiquetado de datos consta de varias tareas, algunas tareas comunes incluyen:

1. **Clasificación de imágenes** : aquí, el objetivo es asignar una etiqueta o categoría a una imagen, como "gato" o "perro".
2. **Detección de objetos** : generalmente se refiere a identificar y ubicar objetos dentro de una imagen, como automóviles o personas.
3. **Segmentación semántica** : Se trata de asignar una etiqueta a cada píxel de una imagen, proporcionando una comprensión detallada del contenido de la imagen. Algunas otras tareas de segmentación implican **la segmentación panóptica** y de instancias .
4. **Reconocimiento de entidades nombradas** : en el reconocimiento de entidades nombradas, el objetivo es identificar y clasificar entidades nombradas, como personas, organizaciones y ubicaciones, dentro del texto.

Estos son solo algunos ejemplos de las muchas tareas que se pueden realizar en el etiquetado de datos.

¿Cuáles son los diferentes tipos de etiquetado de datos?

Hay varios tipos diferentes de etiquetado de datos, y el tipo específico utilizado dependerá de los datos y el uso previsto del conjunto de datos etiquetado. Algunos tipos comunes de etiquetado de datos incluyen:

- **Etiquetado de datos binarios** : Implica asignar una etiqueta de " **verdadero** " o " **falso** " o '1' o '0' a cada dato. Por ejemplo, un conjunto de datos que contiene imágenes de animales se puede etiquetar como "verdadero" si la imagen contiene un gato y "falso" si no lo contiene.
- **Etiquetado de datos multiclas** : generalmente implica asignar etiquetas a más de dos clases en el conjunto de datos dado. Por ejemplo, un conjunto de datos que contiene imágenes de animales puede etiquetarse con los nombres de diferentes especies de animales (p. ej., "gato", "perro", "pájaro", etc.).
- **Etiquetado de datos de múltiples etiquetas** : Se trata de asignar múltiples etiquetas a cada dato. Por ejemplo, un conjunto de datos que contiene imágenes de animales se puede etiquetar con los nombres de los animales en cada imagen, así como el tipo de entorno en el que se encuentran (p. ej., "gato en un campo de hierba"). Este tipo de conjunto de datos es muy popular en la generación de texto a imagen.
- **Etiquetado semántico de datos** : esto implica agregar anotaciones o descripciones detalladas a los datos. Por ejemplo, un conjunto de datos que contenga imágenes de animales podría etiquetarse con descripciones detalladas de la apariencia, el comportamiento y el entorno de los animales.
- **Etiquetado de datos estructurados** : Es etiquetar los datos en un formato estructurado, como una tabla o base de datos. Por ejemplo, un conjunto de datos que contiene comentarios de los clientes puede etiquetarse con los problemas o temas específicos que aborda cada comentario, así como el nombre del cliente y la información de contacto.
- **Etiquetado de datos no estructurados** : Esto implica etiquetar datos que no están organizados en un formato estructurado. Por ejemplo, en NLP y reconocimiento de voz, un conjunto de datos que contiene grabaciones de audio puede etiquetarse con transcripciones de las palabras pronunciadas en cada grabación.

¿Cuáles son los diferentes enfoques para el etiquetado de datos?

Hay dos enfoques:

1. **El etiquetado manual** es el proceso de asignación de etiquetas a los datos por parte de un anotador humano. Esto puede llevar mucho tiempo y puede introducir errores si la persona que etiqueta los datos no tiene cuidado o no está familiarizada con la tarea en cuestión.
2. **El etiquetado automatizado** es el proceso de usar algoritmos y software para asignar automáticamente etiquetas a los datos. Esto puede ser más rápido y más preciso que el etiquetado manual, pero requiere entrenar los algoritmos en un conjunto de datos etiquetado, lo que puede requerir mucho trabajo y mucho tiempo. Además, es posible que el etiquetado automático no siempre sea tan preciso como el etiquetado manual, especialmente para tareas complejas o datos que son difíciles de etiquetar.

Es importante tener en cuenta que el etiquetado automático depende en gran medida de los datos para los que se entrenó el algoritmo. Si la consistencia de los datos entrenados no coincide con la consistencia de los datos sin etiquetar, es posible que el algoritmo no produzca resultados precisos.

Etiquetado manual de datos

Como se mencionó anteriormente, el etiquetado manual de datos involucra a un anotador humano. Este proceso generalmente lo realiza un equipo de anotadores que están capacitados para comprender la tarea en cuestión y las etiquetas específicas que deben asignarse. Los anotadores generalmente revisan cada punto de datos y asignan la etiqueta adecuada en función de su comprensión de los datos, las pautas de etiquetado y la tarea en cuestión.



El etiquetado manual de datos se usa a menudo cuando los datos son complejos o cuando se necesitan etiquetas de alta calidad que requieren interpretación humana y toma de decisiones. Por ejemplo, en tareas relacionadas con la medicina y la asistencia sanitaria. También se puede utilizar cuando no existen algoritmos o modelos que puedan etiquetar con precisión los datos.

¿Cuáles son las ventajas del etiquetado manual de datos?

Una ventaja del etiquetado manual de datos es que permite un mayor control y precisión en el proceso de etiquetado. Debido a que las etiquetas son asignadas por personas capacitadas para comprender los datos y las pautas de etiquetado, a menudo son más precisas y consistentes que las generadas por métodos automáticos. Esto puede ser particularmente importante para tareas que requieren etiquetas de alta calidad, como tareas de clasificación de imágenes o audio.

¿Cuáles son las desventajas del etiquetado manual de datos?

Como desventaja, el etiquetado manual de datos puede llevar mucho tiempo y ser costoso, especialmente para grandes conjuntos de datos. También requiere un equipo de anotadores capacitados, que pueden ser difíciles de encontrar y administrar.



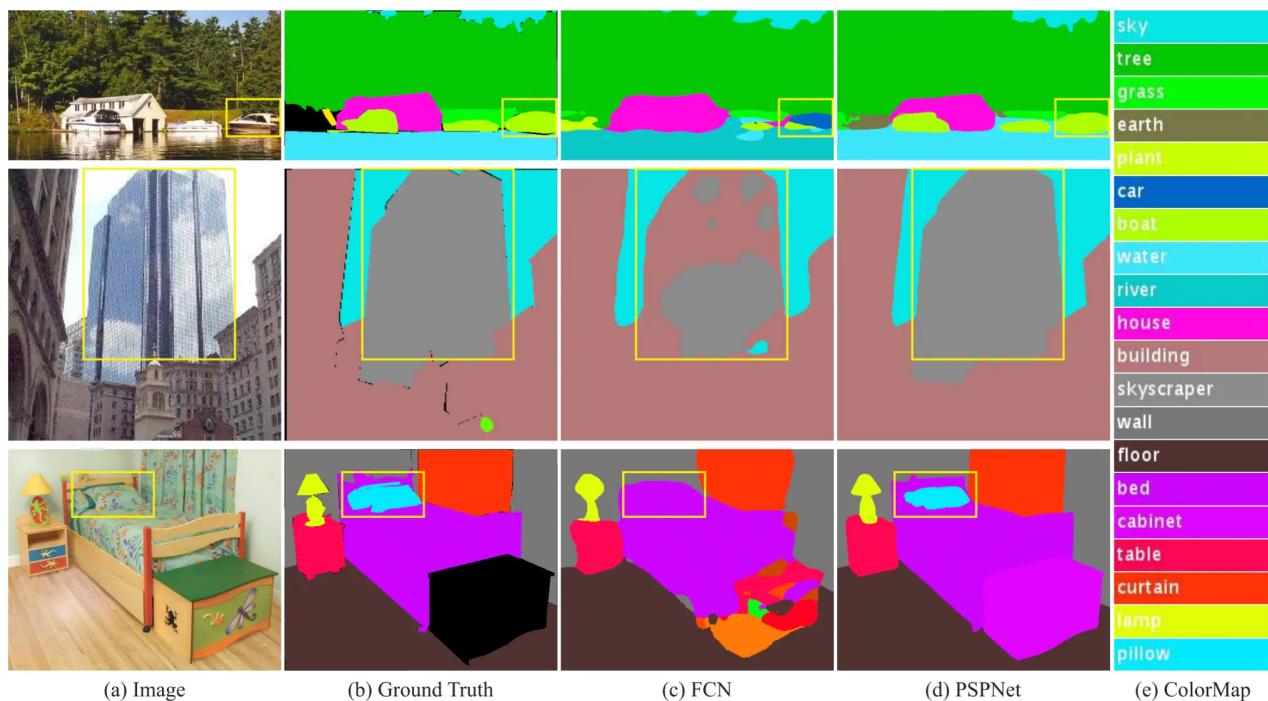
Etiquetado automático de datos

Los procesos de etiquetado automático de datos o la anotación de datos asistida por IA es un proceso en el que se utilizan algoritmos de inteligencia artificial (IA) para ayudar a los anotadores humanos a etiquetar los datos. En cierto sentido, es un etiquetado de datos híbrido o semiautomático. Esto se puede hacer de diferentes maneras, como proporcionando sugerencias para etiquetas basadas en los datos (algo así como un sistema de recomendación), o generando automáticamente etiquetas que pueden ser revisadas y corregidas por anotadores humanos.

La anotación de datos asistida por IA se usa a menudo para mejorar la **eficiencia** y la **precisión** del proceso de etiquetado. Mediante el uso de algoritmos de IA para ayudar en el proceso de etiquetado, los anotadores humanos pueden concentrarse en tareas más complejas o matizadas, mientras que los algoritmos de IA manejan los aspectos más rutinarios o repetitivos del proceso de etiquetado. Esto puede ayudar a reducir la cantidad de tiempo y recursos necesarios para el etiquetado, al mismo tiempo que mejora la calidad de las etiquetas.

¿Cuáles son las ventajas de la anotación de datos asistida por IA?

Una ventaja de la anotación de datos asistida por IA es que puede proporcionar un enfoque equilibrado que combina las fortalezas del aprendizaje humano y automático. Los anotadores humanos pueden proporcionar la experiencia y las habilidades de interpretación necesarias para tareas complejas, mientras que los algoritmos de IA pueden manejar los aspectos más rutinarios del etiquetado, como el reconocimiento de imágenes o la clasificación de texto. Esto puede ayudar a mejorar la precisión y consistencia de las etiquetas, al mismo tiempo que reduce el tiempo y los recursos necesarios para el etiquetado.



Fuente

La imagen de arriba compara el rendimiento de las dos redes neuronales profundas

¿Cuáles son las desventajas de la anotación de datos asistida por IA?

Sin embargo, la anotación de datos asistida por IA también tiene sus limitaciones. Se basa en la disponibilidad de datos de entrenamiento etiquetados para entrenar los algoritmos de IA, que pueden no estar siempre disponibles o pueden ser difíciles de obtener. También requiere el desarrollo y la implementación de algoritmos de IA que puedan ayudar en el proceso de etiquetado, que puede ser un proceso complejo y que requiere mucho tiempo.

Tendencias actuales del etiquetado de datos

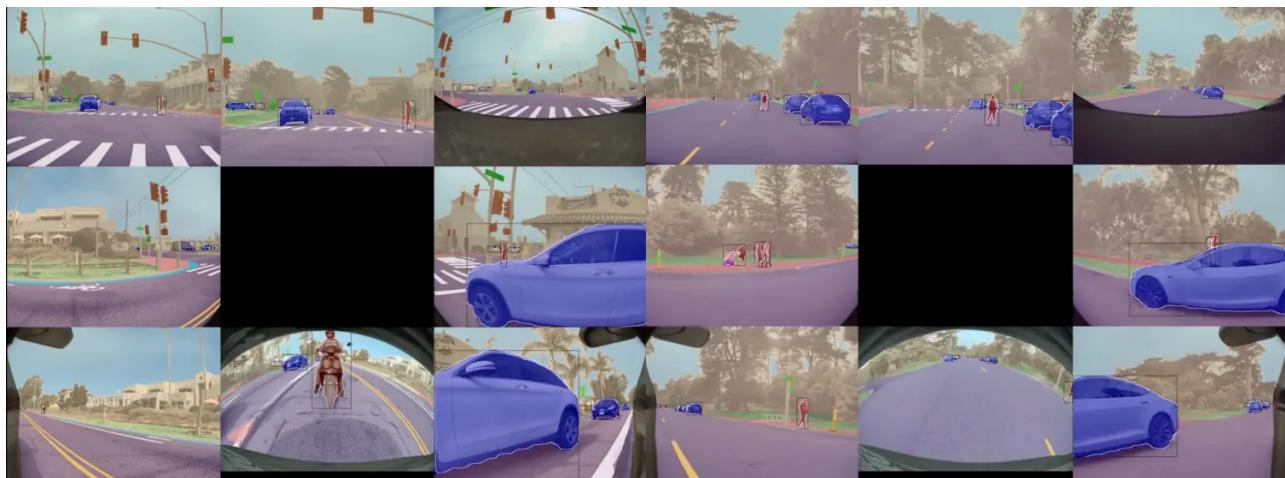
Hay varias tendencias actuales en el etiquetado de datos, que incluyen:

- **El uso de inteligencia artificial y algoritmos de aprendizaje automático para automatizar el proceso de etiquetado :** muchas organizaciones utilizan algoritmos de inteligencia artificial y aprendizaje automático para etiquetar automáticamente sus datos, lo que reduce la necesidad de trabajo manual y acelera el proceso de etiquetado.

- **El desarrollo de nuevas herramientas y tecnologías para el etiquetado de datos :** se están desarrollando muchas herramientas y tecnologías nuevas para el etiquetado de datos, incluidas interfaces fáciles de usar, ayudas visuales y mecanismos de control de calidad. Estas herramientas facilitan el etiquetado de datos de manera precisa y eficiente.
- **Mayor énfasis en la calidad y precisión de las etiquetas de datos :** con la creciente importancia del aprendizaje automático y la IA, hay un mayor énfasis en garantizar que las etiquetas de datos sean precisas y reflejen el verdadero significado de los datos. Esto es importante para entrenar modelos de aprendizaje automático de alta calidad y lograr un buen rendimiento.
- **El uso de crowdsourcing y otros métodos colaborativos para el etiquetado de datos :** muchas organizaciones están utilizando crowdsourcing y otros métodos colaborativos para etiquetar sus datos, lo que les permite aprovechar el conocimiento y la experiencia de una gran comunidad de anotadores.
- **La creciente importancia del etiquetado de datos en una amplia gama de industrias:** el etiquetado de datos se está utilizando en una gama cada vez más diversa de industrias, desde atención médica y finanzas hasta comercio minorista y transporte. A medida que más organizaciones reconocen el valor de los datos etiquetados para el aprendizaje automático y la IA, crece la demanda de servicios de etiquetado de datos.

Estudio de caso: ¿Cómo etiqueta Tesla sus datos?

Tesla utiliza una variedad de [métodos para etiquetar los datos](#) de sus autos autónomos. Emplean un equipo de anotadores humanos que etiquetan manualmente los datos, y también usan algoritmos de inteligencia artificial y aprendizaje automático para etiquetar automáticamente los datos.



Fuente

Para etiquetar los datos, los anotadores de Tesla usan una herramienta especializada que les permite ver los datos (como imágenes o videos capturados por las cámaras del automóvil) y agregar etiquetas o anotaciones. Los anotadores pueden etiquetar objetos en los datos, como otros vehículos, peatones o señales de tránsito, o pueden etiquetar eventos o acciones, como cambios de carril o giros.

Además del etiquetado manual, Tesla también utiliza inteligencia artificial y algoritmos de aprendizaje automático para etiquetar automáticamente los datos. Estos algoritmos se entranen en un gran conjunto de datos etiquetados y luego se usan para etiquetar nuevos datos. Los algoritmos toman la decisión final sobre las etiquetas y anotaciones, sin intervención humana.

En general, el proceso de etiquetado de datos de Tesla está diseñado para producir un conjunto de datos etiquetados que sea preciso, consistente y completo, y que se pueda usar para entrenar sus modelos de automóviles autónomos.

Conclusión

El etiquetado de datos es una tarea importante, ya que permite que los algoritmos de aprendizaje automático produzcan resultados precisos. Los enfoques para el etiquetado de datos pueden variar según la tarea en cuestión. Por ejemplo, el etiquetado manual de datos es un enfoque importante cuando se trata de áreas como la médica y la asistencia sanitaria. Para tareas relacionadas con la secuenciación del genoma o la radiología donde los datos son extremadamente complejos y frágiles, la anotación manual es imprescindible o al menos hasta el punto en que los algoritmos pueden superar a los anotadores humanos. Pero ciertamente, para tareas que involucran vehículos autónomos donde el área de interés es bastante grande, el etiquetado automatizado será suficiente siempre que haya intervención humana.

El etiquetado de datos ciertamente está experimentando un cambio de un enfoque manual a un enfoque automático. Tal vez en los próximos 5 años, la mayoría de las industrias utilizarán procesos de etiquetado automatizados, ya que el ritmo al que evoluciona la IA es bastante seguro.

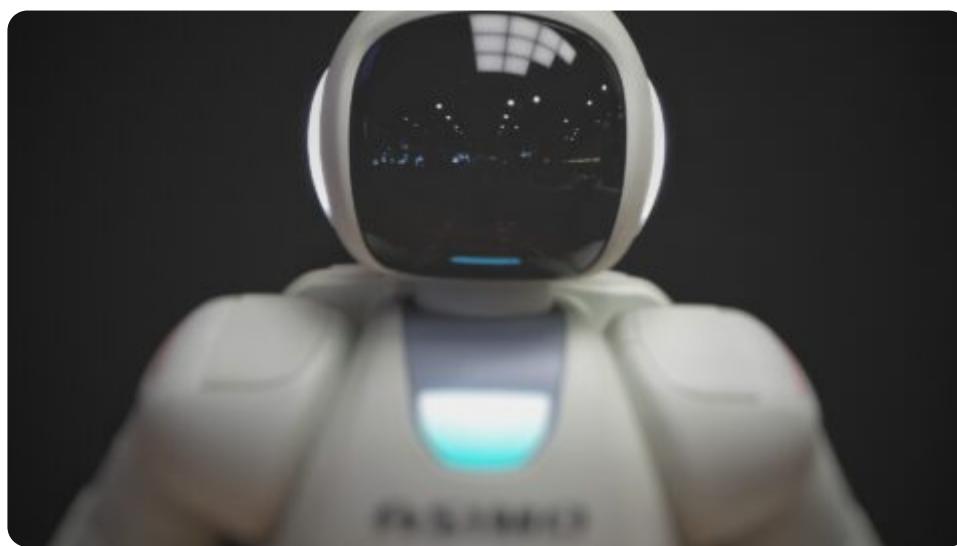


catalán nativo

Artículos relacionados



Los beneficios del etiquetado automático para ML



Libere el poder de los catálogos de datos de aprendizaje automático



Cómo elegir su mejor proveedor de subcontratación de etiquetado de datos

Suscríbase a nuestro boletín de noticias

Correo electrónico *

Entregar

Soluciones

- Etiquetado de visión artificial
- Anotaciones de imagen
- Validación y ajuste del modelo
- Clasificación y encuestas
- Similitud y Comparación
- Validación de datos sintéticos
- Transcripción
- Anotaciones de video
- Moderación de contenido

Industrias

- Drones y Robótica
- Vehículos Autónomos
- Agricultura
- Deportes, Medios y Entretenimiento
- Comercio electrónico y venta minorista

Compañía

- Sobre nosotros
- ¿Por qué Tarea?
- Blog
- Seminarios web y videos
- Glosario
- Preguntas frecuentes
- Carreras
- Preguntas más frecuentes

Estudios de caso

- "Encontrar una pequeña plaga en un campo grande"
- Pago sin fricciones para cadenas minoristas internacionales
- ¿Qué tan rápido se tarda en etiquetar una imagen con 3000 objetos?

Comparar

- Tarea vs Escala
- Tasq vs Appen
- Tasq frente a Labelbox
- Tasq contra Lionbridge
- Tarea vs Competidores

Contacto

- Reserve una demostración
- Contáctenos

[Condiciones de uso](#)

[política de privacidad](#)